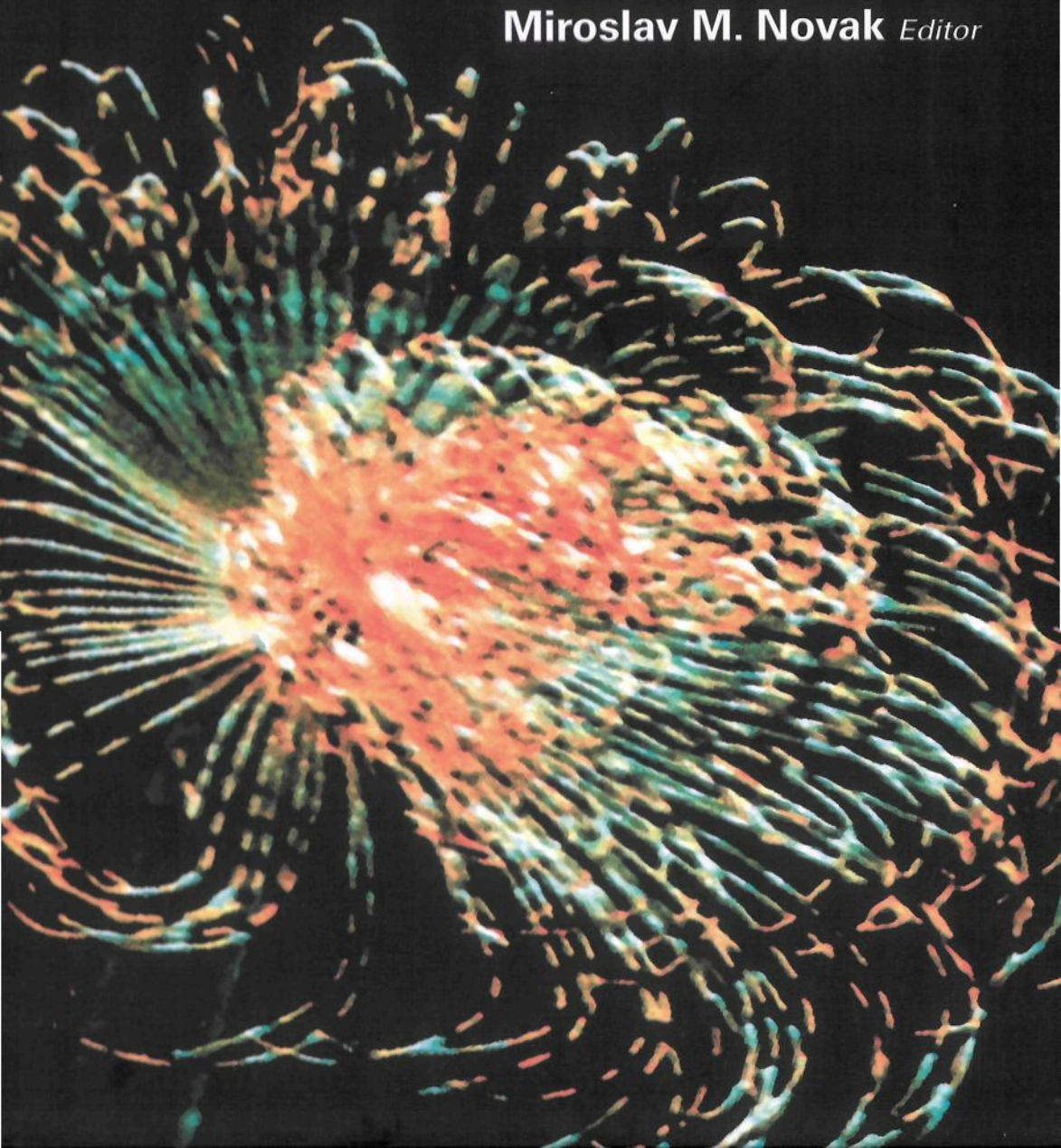# Complexus Mundi
Emergent Patterns in Nature

Miroslav M. Novak *Editor*

Complexus Mundi
Emergent Patterns in Nature

This page is intentionally left blank

Complexus Mundi

Emergent Patterns in Nature

Editor

Miroslav M. Novak

Kingston University, UK

World Scientific

COMPLEXUS MUNDI
Emergent Patterns in Nature

to Erika, Karl and Doreen

This page is intentionally left blank

Contents

This page is intentionally left blank

Preface

The world around us adapts over time in response to changing conditions. New patterns emerge all the time, necessitating the development of theories that underpin the driving processes. Practically all natural processes are complex and natural scientists have began to grapple with the complexity studies to manage and influence the driving forces. The nature of relationships and interactions between humans has, in turn, led the social scientists to adopt ideas from these studies to develop a framework better amenable to measurement and interpretations.

The evolution process is driven by self-organization, and not by some centralized entity. The resulting complexity of interactions makes it impossible to predict, with a degree of certainty, many aspects of system behaviour. The nonlinear dependance between cause and effect lies at the root of these dynamical interactions.

This book offers a unique viewpoint into the complementary aspects of fractals and complexity and provides a unifying thread in this multidisciplinary research. Do the nonlinear interactions play a role in the complexity management of socio-political systems? Is it possible to extract the global properties of genetic regulatory networks without knowing details of individual genes? What can one learn by transplanting the self-organization effects known in laser processes to the study of emotions? What can the change in the level of complexity tell us about the physiological state of the organism? The reader will enjoy finding the answers to these and many other questions in this book.

This volume is the 9th is this series of publications, which are based on presentations at regular conferences organized in different locations worldwide. The distinguishing feature of the series is its vibrant, multidisciplinary nature taking place in collegiate atmosphere. The multidisciplinary feature has been growing more prominent with every new event.

Fractal 2006 was made possible through the help of the following members of the programme committee (in alphabetical order):

P. Allegrini (*Italy*), M. Cieplak (*Poland*), D. Dhar (*India*), K. Falconer (*UK*), L. Gardini (*Italy*), H. Haken (*Germany*), R. Hilfer (*Germany*), A. Holden (*UK*), A. Hubler (*USA*), H. J. Jensen (*UK*), L. S. Liebovitch (*USA*), S. Lowen (*USA*), F. Mainardi (*Italy*), Ch. Mira(*France*), T. Nagatani (*Japan*), E. Scalas (*Italy*), W. H. Steeb (*South Africa*), C. Tricot (*France*), B. J. West (*USA*)

The *Fractal 2006* conference was partially supported by Emergentis Ltd.

The details about the next conference in this series will be posted on the web site http://www.kingston.ac.uk/fractal/.

M. M. Novak
Kingston-upon-Thames, UK

STRUCTURE OF GENETIC REGULATORY NETWORKS: EVIDENCE FOR SCALE FREE NETWORKS

L. S. LIEBOVITCH

Florida Atlantic University
Center for Complex Systems & Brain Sciences
Center for Molecular Biology & Biotechnology
Department of Psychology
Department of Biomedical Sciences
777 Glades Road, Boca Raton FL, U.S.A. 33431
E-mail: liebovitch@ccs.fau.edu

V. K. JIRSA

Florida Atlantic University
Center for Complex Systems & Brain Sciences
Department of Physics
777 Glades Road, Boca Raton FL, U.S.A. 33431
E-mail: jirsa@ccs.fau.edu

L. A. SHEHADEH

Florida Atlantic University
Center for Complex Systems & Brain Sciences
Department of Biomedical Sciences
777 Glades Road, Boca Raton FL, U.S.A. 33431
E-mail: shehadeh@ccs.fau.edu

The expression of some genes increases or decreases the expression of other genes forming a complex network of interactions. Typically, correlations between the expression of different genes under different conditions have been used to identify specific regulatory links between specific genes. Instead of that "bottom up" approach, here we try to identify the global types of networks present from the global statistical properties of the mRNA expression. We do this by comparing the statistics of mRNA computed from different network models, including random and fractal, scale free networks, to that found experimentally. The novel features of our approach are that: 1) we derive an explicit form of the connection matrix between genes to represent scale free models with arbitrary scaling exponents, 2) we extend Boolean networks to quantitative regulatory connections and quantitative mRNA expression concentrations, 3) we identify possible types of global genetic regulatory networks and their parameters from the experimental data.

1

1. Introduction

The information encoded in deoxyribonucleic acid, DNA, in genes is transcribed into messenger ribonucleic acid, mRNA, that is then translated into proteins that perform the structural and functional characteristics of cells. Some of these proteins, called transcription factors, then bind back onto DNA increasing, or decreasing, the expression of other genes. Recent co-regulational studies have used different computational methods to evaluate the correlations in mRNA concentrations under different experimental conditions to determine the network of genetic regulation. These methods include correlation analysis, cluster analysis, neural networks, genetic algorithms, and singular value decomposition (principal component analysis) with or without additional constraints [1]-[18]. Instead of the "bottom up" method of those co-regulational studies, our approach here emphasizes a "top down" method to try to identify global patterns of genetic regulation. We do this in the following way: 1) We formulate different types of models of networks of genetic regulation. 2) We compute the probability density function, PDF, of the mRNA from each network. 3) We then compare those results to the experimental data from mRNA concentrations measured by cDNA microarray chips. By looking for the global features of the genetic regulatory network, as characterized by the PDF of the mRNA, we may be able to identify different types of genetic regulatory networks without the need to first identify which specific gene regulates which other specific gene.

2. Models of Genetic Regulatory Networks

We represent the genetic regulatory network of N genes by a connection matrix M, where $M(i, j)$ is the regulatory effect of the j-th gene on the i-th gene. Thus, the regulatory effects can be made quantitative which is an extension of the all or nothing (0 or 1) of Boolean models. Connection matrices with both positive (stimulatory) and negative (inhibitory) elements can cause a broad range of complex behavior. We are therefore faced with a trade off between studying less complex models that we can successfully analyze and more complex models which would be much more difficult to analyze. Therefore, in these initial studies, we restrict ourselves to connection matrices with only stimulatory behavior where all the elements are greater than or equal to zero so that the mRNA concentrations will always reach a steady state with a characteristic probability density function.

We studied connection matrices with random, [19] fractal scale free, and small world [20],[21],[22] connection topologies. Each connection matrix was 1000 x 1000 to model the interactions between 1000 genes. Interactions between genes i and j were represented by $M(i, j) = 1$, while all other $M(i, j) = 0$.

Each different model is characterized by the degree distribution of the number of genes g connected to other genes k. In the models with random connections, the degree distribution, of either the inputs into the genes or the outputs from the genes, have the Poisson distribution $g(k) = \frac{<k>^k}{k!} e^{-<k>}$ where there are on average $< k >$ interactions between genes. We formulate both Random Output and Random Input

models.

In the fractal scale free models the degree distribution has the power law form $g(k) = Ak^{-a}$. We formulate three types of such models. First, where the degree distribution of the outputs from the genes is a such a power law. In this model the statistical pattern of input regulation into each gene is the same, namely each gene has a small number of inputs from the lower order genes and ever more inputs from ever more higher order genes. We call this the Scale Free Homogeneous model since the statistical pattern of inputs into each gene is the same. In the second model, the statistical pattern of input regulation into each gene is a power law. Different genes have different average numbers of connections to other genes. We therefore call this model the Scale Free Heterogeneous model. In the third model, we adjust both the input and output degree distributions to have the same form, and therefore call this model the Scale Free Symmetric model.

To formulate the scale free models we chose the number of nonzero elements in the columns of $M(i,j)$ for the Homogeneous model, the number of nonzero elements in the rows of $M(i,j)$ for the Heterogeneous model, and the number of nonzero elements in diagonals of $M(i,j)$ for the Symmetric model as given by $Y(r) = (\frac{A}{(a-1)(N-r)})^{1/(a-1)}$ where r is the distance, respectively, from the starting column, row, or corner of the matrix. This relationship was derived by noting that the number $g(k)dk$ of genes with k connections over the interval $(k, k+dk)$ is equal to the number of nonzero elements, dr, with $Y(r) = k$ connections over the corresponding interval $(r, r+dr)$. This relationship between $g(k)$ and $Y(r)$ provides a useful explicit way to generate connection matrices corresponding to a scale free model with a given scaling exponent a without having to from the network iteratively using a growth rule to preferentially add nodes [23,24]. In all three models we varied the scaling parameter over the range $1 \leq a \leq 6$. We also computed the average path length between any pairs of genes through their connections and the clustering coefficient [29]. All of the scale free networks had both shorter path length and higher clustering coefficient than that of the random networks with the same $< k >$ indicating that these scale free networks were all "small world" networks.

In both the random and scale free models we adjusted the density of non-zero elements so that each model had the same number of average connections $< k >= 6$ between genes.

3. Statistics of the mRNA from the Different Models

We then represent the mRNA level expressed by gene i of the N genes by an element in the column vector $X(i)$. We start with initial values of the elements of $X(i)$. At each subsequent time step the new elements of $X'(i)$ are then computed from $X' = MX$.

Some of the models presented here can be cast in a the form of Markovian matrices, that is, where each row (or column) has an identical sum (equal to 1 if all the elements are properly normalized). In that case, the matrix M has an eigenvalue

of unity and the iteration process $X' = MX$ converges to a steady state. However, the scale free models cannot be cast into that form. In that case, we renormalized the norm of the vector X at each time step. That is, for these models the relative values of the elements reach a steady state in the direction of the vector X, although its length continues to increase or decrease. In reality, biological constraints, such as the finite number of mRNA molecules that can be synthesized, will constrain the norm of the vector X to a finite nonzero value.

Some models, for example, the Random Output model, can be formulated as either a Markovian or non-Markovian matrix. For such models we found that the direct iteration of $X' = MX$ of the Markovian matrix and the renormalization of X at each time step of the iteration of non-Markovian matrix both produced very similar steady state statistical patterns of mRNA expression, giving credence to our renormalization procedure used to compute the models with non-Markovian matrices.
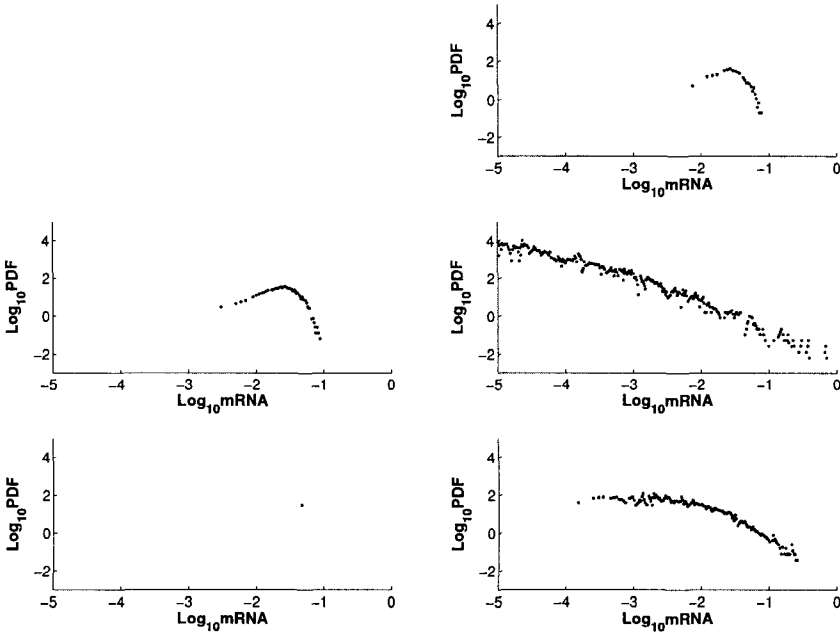
Figure 1. PDFs of mRNA concentrations computed from different models of genetic regulatory networks. Counter-clockwise starting from the top left: 1) Random Output model, 2) Random Input model, 3) Scale Free Symmetric model with scaling exponent a=2, 4) Scale Free Heterogeneous model with a=2, and 5) Scale Free Homogeneous model with a=-2.

At steady state, we then determine the PDF, the probability density function

that any of the mRNA concentrations, $X(i) = x$, is between x and $x + dx$. Our approach is to analyze the statistical form of the PDF of all the mRNA concentrations, and we do not determine which specific gene is associated with which specific mRNA concentration. That makes it possible for our approach to identify global properties of the genetic regulatory network without first needing to identify the specific regulatory connections between specific genes.

The standard method is to compute the PDF from the histogram of the number of values $n(x)$ in each interval $[x, x + dx]$, namely $PDF(x) = \frac{n(x)}{N dx}$. Using this method, when dx is small, the PDF is well characterized at small x, but poorly characterized at large x since there are few values in each narrow bin in the tail of the distribution. When dx is large more values are captured in each bin at large x, but now the resolution of the PDF is poor at small x. We overcome these limitations by using an algorithm that combines histograms of different bins size dx so that it generates PDFs that are accurate both at small dx and at large dx [25]. The results for 2 random and 3 scale free models are shown in Fig. 1

4. Experimental Data

The mRNA concentrations expressed by thousands or tens of thousands of genes can now be measured simultaneously by cDNA microarray chips where flourescently marked mRNA binds to complementary polynucleotide sequences bound to the chip [26,27]. We calculated the PDFs of the mRNA concentrations measured by cDNA microarray chips used to study circadian rhythms in Drosophila flies [28]. Figure 2 shows the PDF determined from wild type flies kept in total darkness. This form
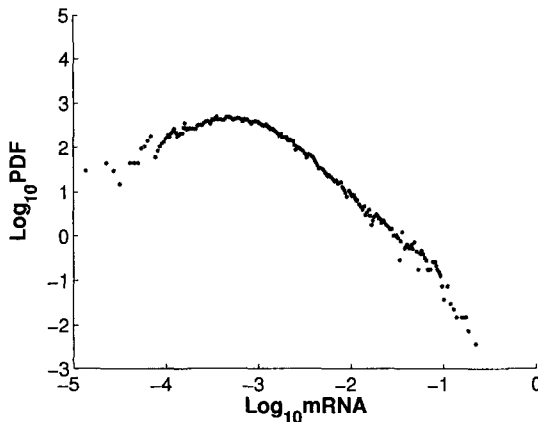
Figure 2. Typical PDF of mRNA concentrations computed from experimental data of virtually all the genes in Drospholia.

of the PDF is typical of several different light and dark conditions for both the wild

type flies and those with a mutation in the period gene that alters their circadian rhythm.

5. Conclusions

The experimental data analyzed here is most like the genetic regulatory network model of a fractal, Scale Free Symmetric network with a scaling exponent a of approximately 2 for the distributions of both the inputs into the genes and the outputs from the genes. However, since we have not shown that the PDFs of the mRNA concentrations are unique for each model, it is possible that other models that we have not studied might also provide an equally good fit to the experimental data.

The Scale Free Symmetric model means that some genes are strongly regulated by the expression of only a few other individual genes, while other genes are regulated by the average expression of a large number of genes. Moreover, the degree distribution of the input regulation into each gene is the same as the degree distribution of the output regulation from each gene.

The strength of our approach presented here is that we have been able to extract some information about the global properties of the genetic regulatory network without first determining which specific genes regulate which other specific genes. This was possible because we compared the PDF of the mRNA predictions of the models to the experimental data without considering which particular gene produced which particular mRNA concentration.

Since we have shown here that different network architectures may have different PDFs of the mRNA concentrations, this may provide a way to identify different subnetworks, for example, metabolic vs. cell cycle regulatory networks. It may also prove useful in determining if those networks change under different conditions, for example, normal vs. metastatically transformed cells. If that is the case, then changes in the PDFs may be useful in identifying such cellular transformations and the degree of those transformations. If the PDFs are not different under different conditions, it may indicate that the overall global structure of these genetic regulatory networks is so important to cell function that as some genes are turned off, other genes may be recruited to fulfill their missing function in the entire network, so that the overall structure of the network remains intact.

Some aspects of these studies may also be useful in the analysis of other network systems. For example, we derived the distribution of nonzero elements in the connection matrix corresponding to scale free distributions with different scaling exponents. This makes it possible to explicitly construct scale free networks without the need to form them iteratively by using growth rules to add nodes to the network. An important question about networks is the relationship between their structure and dynamics. For example, in many experimental cases only the activity of the nodes of the network can be measured and one seeks to use that information to determine the structure of the network. For example, there are electrical record-

ings from nerve cells and one seeks to use this information to understand how they are functionally connected, or there are traffic activity measures at routers and one seeks to understand the topology of the Internet. We found here that the tails of the PDFs of the mRNA concentrations are related to the degree distribution. The mRNA concentrations correspond to the dynamic activity at the nodes of a network and the degree distribution to the structure of a network. This relationship forms an interesting link between the dynamics and the structure of a network, that may be, or may not be, present in other types of networks as well.

6. Acknowledgments

References

1. J.L DeRisi, V.R. Iver, and P.O. Brown, *Science*, **278**, 680-6 (1997). (Data online at http://cmgm.stanford.edu/pbrown/explore/index.html).
2. M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein, *Proc Natl Acad Sci USA*, **95**, 14863-8 (1998).
3. G.S. Michaels, D.B. Carr, M. Askenazi, S. Fuhrman, X. Wen, R. Somogyi, *Pac Symp Biocomput* ,**3**, 42-53 (1998).
4. P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown , D. Botstein, B. Futcher, *Mol. Biol. Cell*, **9**, 3273-97 (1998). (Data online at http://cellcycle-www.stanford.edu/).
5. X. Wen, S. Fuhrman, G.S. Michaels, D.B. Carr, S. Smith, J.L. Barker, and R. Somogyi, *Proc. Natl. Acad. Sci. USA*, **95**, 334-9 (1998).
6. M.K.S Yeung, J. Tegner, and J.J. Collins, *Proc. Natl. Acad. Sci. USA*, **9**, 6163-68 (1999).
7. A.J. Hartermink, D.K. Gifford, T.S. Jaakkola,and R.A. Young, *Symp. Biocomputing*, **6**, 422-33 (2001).
8. P. D'haeseleer, X. Wen, S. Fuhrman, R. Somogyi, *Pac. Symp. Biocomputing*, **4**, 41-52 (1999).
9. S. Tavazoie, J.D. Hughes, M.J. Campbell, R.J. Cho, and G.M. Church, *Nat. Genet*, **22**, 281-5 (1999).
10. p. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, and T.R. Golub, *Proc. Natl. Acad. Sci. USA*, **96**, 2907-12 (1999).
11. M. Wahde and J. Hertz, *Biosystems*, **55**, 129-36 (2000).
12. P. D'haeseleer, S. Liang, and R. Somogyi, *Bioinformatics*, **16**, 707-26 (2000).
13. S. Raychaudhuri, J.M. Stuart, and R.B. Altman, *Pac Symp Biocomput*, **5**, 452-63 (2000).
14. M.P. Brown, W.N. Grundy, D. Lin, N. Cristianini, C.W. Sugnet, T.S. Furey, M Jr. Ares, and D. Haussler, *Proc. Natl. Acad. Sci. USA*, **97**, 262-7 (2000).
15. O. Alter, P.O. Brown, and D. Botstein, *Proc. Natl. Acad. Sci. USA*, **97**, 10101-6 (2000).
16. N.S. Holter, A. Maritian, M. Cieplak, N.V. Fedoroff, and J.R. Banavar, *Proc. Natl. Acad. Sci. USA*, **98**, 1693-8 (2001).

17. T.G. Dewey, and D.J. Galas, *Funct. Integer Genomics*,1, 269-78 (2001).

18. A. Bhan, D.J. Galas, and T.J. Dewey, *Bioinformatics*, **18**, **11**, 1486-93 (2002).

19. P.P. Erdos, and A. Renyi, *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17-61 (1960).

20. A.L. Barabasi, and R.A. Albert, *Science* , **286**, 509-512 (1999).

21. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, *Nature*, **407**, 651-4 (2000).

22. D.J. Watts and S.H Strogatz , *Nature* , **393**, 440-42 (1998).

23. B.A. Huberman and A. L. Adamic, *Nature* **401**, 131-6 (1999).

24. R. Albert and A.L. Barabasi,*Phys. Rev. Lett.* **85**, 5234-7 (2000)

25. L.S. Liebovitch, A.T. Todorov, M. Zochowski, D. Scheurle, L. Colgin, M.A. Wood, K.A. Ellenbogen, J.M. Herre, and R.C. Berstein, *Phys. Rev. E*, **59**, 3312-19 (1999).

26. M. Schena, Ed., DNA Microarrays: A Practical Approach, Oxford Univ. Press, NY (1999).

27. J.B. Rampal, Ed., DNA Arrays: Methods and Protocols, Human Press (2001).

28. Y. Lin, M. Han, B. Shimada, L. Wang, T.M. Gibler, A. Amarakone, T.A. Awad, G.D. Stormo, R.N. Van Gelder, P.H. Taghert, *Proc. Natl. Acad. Sci. USA*, **99**, 9562-7 (2002).

29. R. Albert, and A.L. Barabasi, *Revs. Mod. Phys.*, **74**, 47-97 (2002).

MODELING FRACTAL DYNAMICS

BRUCE J. WEST

Mathematical & Information Sciences Directorate, US Army Research Office,

Research Triangle Park, NC 27709, USA

The proper methodology for describing the dynamics of certain complex phenomena and fractal time series is the fractional calculus through the fractional Langevin equation discussed herein and applied in a biomedical context. We show that a fractional operator (derivative or integral) acting on a fractal function, yields another fractal function, allowing us to construct a fractional Langevin equation to describe the evolution of a fractal statistical process, for example, human gait and cerebral blood flow. The goal of this talk is to make clear how certain complex phenomena, such as those that are abundantly present in human physiology, can be faithfully described using dynamical models involving fractional differential stochastic equations. These models are tested against existing data sets and shown to describe time series from complex physiologic phenomena quite well.

1. Introduction

In the late nineteenth century, most mathematicians felt that a continuous function must have a derivative "almost everywhere", which means the derivative of a function is singular only on a set of points whose total length (measure) vanishes. However, some mathematicians wondered if functions existed that were continuous, but did not have a derivative at any point (continuous everywhere but differentiable nowhere). The motivation for considering such pathological functions was initiated within mathematics, for example, Karl Weierstrass (1815-1897) gave a lecture to the Berlin Academy in 1872 in which he presented functions that had the aforementioned continuity and non-differentiability properties. Twenty-six years later, Ludwig Boltzmann, who elucidated the microscopic basis of entropy, said that physicists could have invented such functions in order to treat collisions among molecules in gases and fluids. It was assumed in the kinetic theory of gases that molecules are materially unchanged as a result of interactions with other molecules, and collisions are instantaneous events as would occur if the

molecules w ere impenetrable and perfectly elastic. A s a result, i t seemed quite natural that the trajectories of molecules would sometimes undergo discontinuous changes. Robert Brown, in 1827, observed the random motion of a speck of pollen immersed in a water droplet. Discontinuous changes in the speed and direction of the motion of the pollen mote were observed, but the mechanism causing these changes was not understood.

Albert Einstein published a paper in 1905 that, although concerned with diffusion in physical systems, ultimately explained the source of Brownian motion as being d ue t o t he net imbalance o f t he r andom c ollisions of t he lighter p articles o f t he medium with the surface of the pollen mote. Jean Baptiste Perrin, of the University of Paris, experimentally verified Einstein's predictions and received the Nobel Prize for his work in 1926. Perrin[1], giving a physicist's view of mathematics in 1913, stated that curves without derivatives are more common than those special, but interesting ones, like the circle, that have derivatives. In fact he was quite adamant in his arguments emphasizing t he i mportance of n on-analytic functions for d escribing c omplex p hysical phenomena, such as Brownian motion. Thus, there are valid physical reasons for looking for these types of functions, but the scientific reasons became evident to the general scientific community only long after the mathematical discoveries made by Weierstrass.

On the theoretical physics side, the Kolmogorov-Arnold-Moser (KAM) theory for conservative dynamical systems, describes how the continuous trajectories of a particle break up into a chaotic sea of randomly disconnected points. Furthermore, the strange attractors of dissipative dynamical systems have a fractal dimension in phase space. Both these developments in classical dynamics, KAM theory and strange attractors, emphasize the importance of non-analytic functions in the description of the evolution of deterministic nonlinear dynamical systems. We do not discuss the details of such dynamical systems herein, but refer the reader to a number of excellent books on the subject ranging from the mathematical rigorous, but readable[2], to provocative picture books[3], to extensive applications[4].

The separation of time scales in physical phenomena allows us to smooth over the microscopic fluctuations and construct a differentiable representation of the dynamics on large space scales and long time scales. However, such smoothing is not always possible, examples of physical phenomena that resist this approach include turbulent fluid flow[5], the stress relaxation of viscoelastic materials such as plastics and rubber[6,7] and finally phase transitions[8,9]. Metaphorically, these complex phenomena, whose evolution cannot be described by ordinary differential equations of motion; leap and jump in unexpected ways to obtain food; they unpredictably twist and turn to avoid capture, and suddenly change strategy to anticipate environmental changes. To understand these and other analogous processes in physiology we find that we must adopt a new type of modeling, one that is not in terms of ordinary or partial differential equations of motion. It is clear that the fundamental elements of complex physical phenomena, such as phase transitions, the deformation of plastics and the stress relaxation of polymers, satisfy Newton's laws. In these phenomena the evolution of individual particles are described by

ordinary differential equations that control the dynamics of individual particle trajectories. It is equally clear that the connection between the fundamental laws of motion controlling the individual particle dynamics and the observed large-scale dynamics cannot be made in any straightforward way.

Herein we describe some of the essential features of fractal functions starting from the simple dynamical processes described by functions that are fractal, such as the Weierstrass function, that are continuous everywhere but are nowhere differentiable. This idea of non-differentiability leads to the introduction of the elementary definitions of fractional integrals and fractional derivatives starting from the limits of appropriately defined sums. We find that the relation between fractal functions and the fractional calculus is a deep one. For example, the fractional derivative of a regular function yields a fractal function of dimension determined by the order of the fractional derivative. Thus, the changes in time of phenomena that are best described by fractal functions are probably best described by fractional equations of motion, as well. In any event, this latter perspective is the one we developed elsewhere[10] and discuss herein. Others have also made inquires along these lines[11].

2. Fractional calculus

It is useful to have in mind the formalism of the fractional calculus before embarking on the interpretation of models using this formalism to explain the complexity of physiological phenomena. What we call the fractional calculus dates back to a question L'Hospital asked Leibniz in 1695, where in a letter he asked the meaning of the expression $d^n y / dx^n$ if $n=1/2$, that is: ..what if n is fractional?

Leibniz replied in part[12]

Thus it follows that $d^{1/2}x$ will be equal to $2\sqrt{dx:x}$ ......John Bernoulli seems to have told you of my having mentioned to him a marvelous analogy which makes it possible to say in a way the successive differentials are in geometric progression. One can ask what would be a differential having as its exponent a fraction. You see that the result can be expressed by an infinite series. Although this seems removed from Geometry, which does not yet know of such fractional exponents, it appears that one day these paradoxes will yield useful consequences, since there is hardly a paradox without utility.

After 310 years of sporadic development, the fractional calculus is now becoming sufficiently well developed and well known that books and articles are being devoted to its consequences in the physical sciences[13,14,15].

One way to introduce fractional operators is by generalizing Cauchy's formula for a n-fold integration over a fixed time interval (a,t)

$$\frac{1}{(n-1)!}\int_a^t (t-\xi)^{n-1} f(\xi)d\xi = \int_a^t \int_a^{t_1} \cdots \int_a^{t_{n-1}} f(\xi_n)d\xi_n \cdots d\xi_1 \equiv {}_aD_t^{(-n)}[f(t)],$$

where the operator ${}_aD_t^{(-n)}[\cdot]$ denotes the n-fold integration operation. The fractional integral analogue to this equation is defined as

$${}_aD_t^{(-\alpha)}[f(t)] = \frac{1}{\Gamma(\alpha)}\int_a^t (t-\xi)^{\alpha-1} f(\xi)d\xi ; \quad t \geq a \tag{1}$$

where the factorial has been replaced by the gamma function, which has an analytic continuation into the complex domain for non-integer and non-positive values of its argument. The fractional derivative corresponding to the fractional integral in (1) is given by

$${}_aD_t^{(\alpha)} = \frac{d^n}{dt^n}{}_aD_t^{(\alpha-n)} \tag{2}$$

where $[\alpha]+1 \geq n \geq [\alpha]$ and the bracket denotes the integer value n closest to α. Consequently for $\alpha < 1/2$ we have $n = 0$. Equation (1) is the Riemann-Liouville ($\mathcal{RL}$) formula for the fractional operator; it is the integral operator when $\alpha < 0$ and it is the differential operator interpreted as (2) when $\alpha > 0$.

2.1 Derivative of a fractal function

Richardson, in his 1926 investigation of turbulence, observed that the velocity field of the atmospheric wind is so erratic that it probably cannot be described by an analytic function[16]. He suggested a Weierstrass function as a candidate to represent the velocity field, since the function is continuous everywhere, but nowhere differentiable, properties he observed in the wind-field data. Here we investigate a generalization of the Weierstrass function in order to simplify some of the discussion

$$W(t) = \sum_{n=-\infty}^{\infty} \frac{1}{a^n}\left[1 - \cos(b^n t)\right] \tag{3}$$

under the conditions that $b > a > 1$. Note that the time derivative of (3) yields a series in which the successive terms increase by a factor of $b/a > 1$ and therefore diverges. The generalized Weierstrass function (GWF) satisfies the scaling relation

$$W(bt) = aW(t), \tag{4}$$

resulting from a simple shift of the index n in the summation. Equation (4) has the form of a renormalization group scaling relation[5], which can be solved by assuming a solution of the form

$$W(t) = \sum_{n=-\infty}^{\infty} A_n t^{H_n} \tag{5}$$

Inserting (5) into (3) yields the equation for the scaling index

$$H_n = \frac{\ln a}{\ln b} - i \frac{2\pi n}{\ln b}, \tag{6}$$

where the scaling exponent is seen to be complex. This exponent has been related to a complex fractal dimension in the architecture of the human lung[17], in many other physiological systems[4,5], as well as in earthquakes, turbulence and financial crashes[18].

The GWF is a superposition of harmonic terms having increasing frequencies as powers of b with decreasing amplitudes as powers of $1/a$. This function has a fractal dimension D if we choose $a = b^{2-D}$, so that in terms of the fractal dimension we write the GWF as

$$W(t) = \sum_{n=-\infty}^{\infty} \frac{1}{b^{(2-D)n}} [1 - \cos(b^n t)]. \tag{7}$$

The $\mathfrak{RL}$-fractional integral of the GWF is given by

$$W^{(-\alpha)}(t) \equiv {}_{-\infty}D_t^{(-\alpha)}[W(t)] = \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{t} \frac{W(\xi)}{(t-\xi)^{1-\alpha}} d\xi \tag{8}$$

for $0 < \alpha < 1$, which after some not so straightforward analysis[4,19] yields

$$W^{(-\alpha)}(t) = \sum_{n=-\infty}^{\infty} \frac{1}{b^{(2-D+\alpha)n}} [1 - \cos(b^n t)] \tag{9}$$

Similarly the $\mathfrak{RL}$-fractional derivative of the GWF is given by

$$W^{(\alpha)}(t) \equiv {}_{-\infty}D_t^{(\alpha)}[W(t)] = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_{-\infty}^{t} \frac{W(\xi)}{(t-\xi)^{\alpha}} d\xi \tag{10}$$

for $0 < \alpha < 1$, which integrates to[4,20]

$$W^{(\alpha)}(t) = \sum_{n=-\infty}^{\infty} \frac{1}{b^{(2-D-\alpha)n}} [1 - \cos(b^n t)]. \tag{11}$$

Consequently, we see that the fractional integral shifts the fractal dimension $D \to D - \alpha$ and the fractional derivative shifts the fractal dimension $D \to D + \alpha$.

These results can be interpreted by noticing that the fractional dimension gives information about the degree of irregularity of the function under analysis. Carrying out a fractional integral of the GWF implies decreasing its fractional dimension and therefore smooths the process, whereas carrying out the fractional derivative means increasing the fractional dimension and therefore making the process and its increments more irregular. What is most intriguing is the fact that a fractional operator acting on a fractal function yields another fractal function, the derivative does not diverge, as does an ordinary derivative of a fractal function, like that of Weierstrass. This suggests that the fractional calculus is the appropriate method for characterizing the dynamics of complex phenomena, particularly those that are described by fractal functions.

2.2 Fractional Brownian motion

It is possible to extend our discussion of GWF to random processes by including random phases in the definition

$$W(t) = \mathrm{Re}\left\{ \sum_{n=-\infty}^{\infty} \frac{1}{a^n} \left[1 - e^{ib^n t} \right] e^{i\phi_n} \right\} \tag{12}$$

where the phase is a random quantity uniformly distributed on the interval $(0, 2\pi)$. Equation (12) would be one way to introduce a random function that has the desired scaling properties and we could discuss the dynamics of this process in time using the fractional operators. An alternative approach is to start with continuous functions and that is what we do now. Consider a function defined by the Fourier transform

$$\xi(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} \hat{\xi}(\omega) d\omega . \tag{13}$$

Operating on this function with the $\mathcal{RL}$-fractional operator, with a lower limit of negative infinity, defines a new function

$$F_\alpha(t) = {}_{-\infty}D_t^{(-\alpha)}[\xi(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} (-i\omega)^{-\alpha} \hat{\xi}(\omega) d\omega \tag{14}$$

and the weighting in the integrand is obtained by operating on the exponential in the Fourier transform ${}_{-\infty}D_t^{(-\alpha)}\left[e^{i\omega t}\right] = e^{i\omega t} (i\omega)^{-\alpha}$. If we now interpret the function (13) as a stochastic quantity, then we can evaluate the correlation of the function (14) at two time points separated by a time interval τ

$$\left\langle F_\alpha(t)F_\alpha^*(t+\tau)\right\rangle_\xi = \frac{1}{(2\pi)^2}\int\limits_{-\infty}^{\infty}d\omega_1\int\limits_{-\infty}^{\infty}d\omega_2 e^{-i(\omega_1-\omega_2)t}e^{i\omega_2\tau}\omega_1^{-\alpha}\omega_2^{-\alpha}\left\langle\hat{\xi}(\omega_1)\hat{\xi}^*(\omega_2)\right\rangle_\xi \quad (15)$$

where the brackets denote an average over an ensemble of realizations of random fluctuations. If the random fluctuations correspond to a Wiener process the average in the integrand reduces to

$$\left\langle\hat{\xi}(\omega_1)\hat{\xi}^*(\omega_2)\right\rangle_\xi = C\delta(\omega_1-\omega_2) \quad (16)$$

and C is a constant. Substituting (16) into (15) and integrating over one of the frequencies yields for the autocorrelation function

$$\left\langle F_\alpha(t)F_\alpha^*(t+\tau)\right\rangle_\xi = \frac{C}{(2\pi)^2}\int\limits_{-\infty}^{\infty}d\omega e^{i\omega\tau}\omega^{-2\alpha} \propto \tau^{2\alpha-1}. \quad (17)$$

If we make the association of the order of the fractional operator with the Hurst exponent

$$\alpha = H - 1/2 \quad (18)$$

we observe that the fractional index lies in the interval $-1/2 \le \alpha \le 1/2$, because the Hurst exponent is confined to the interval $0 < H \le 1$. Therefore the solution to the fractal stochastic equation of motion

$$_{-\infty}D_t^{(\alpha)}[F_\alpha(t)] = \xi(t) \quad (19)$$

given by (14) is a colored noise representation of the dynamics expressed by (19). The integral relation in (14) is linear so that the solution $F_\alpha(t)$ and the random fluctuations $\xi(t)$ have the same statistics, which, by assumption, are Gaussian. Consequently the solution $F_\alpha(t)$ is a realization of fractional Brownian motion with the fractional index restricted to the indicated region.

3. Fractional Langevin equations

Of course, the fractional calculus does not in itself constitute a physical/biological theory, but requires such a theory in order to interpret the fractional derivatives and integrals in terms of physical/biological phenomena. We therefore follow a pedagogical approach and examine the simple relaxation process described by the rate equation

$$\frac{d}{dt}\Phi(t) + \lambda\Phi(t) = 0 \quad (20)$$

where $t > 0$ and the relaxation rate λ determines how quickly the process returns to its equilibrium state. The solution to (20) is given by $\Phi(t) = \Phi(0)e^{-\lambda\tau}$ which is unique in

terms of the initial condition $\Phi(0)$. An alternative way of writing (20) is in terms of the anti-derivative operator

$$\Phi(t) - \Phi(0) = \lambda \left(\frac{d}{dt} \right)^{-1} \Phi(t)$$

which suggests, for its generalization, replacing the anti-derivative with the $\mathcal{RL}$ - fractional integral operator

$$\Phi(t) - \Phi(0) = \lambda^{\alpha} {}_0 D_t^{(-\alpha)} [\Phi(t)] \tag{21}$$

where the lower limit of the fractional integral is zero, corresponding to the initial value problem. Operating on the left in (21) with the fractional derivative we obtain the generalization to the relaxation equation given by[21]

$${}_0 D_t^{(\alpha)} [\Phi(t)] + \lambda^{\alpha} \Phi(t) = \frac{t^{-\alpha} \Phi(0)}{\Gamma(1 - \alpha)} \tag{22}$$

and the initial value becomes an inhomogeneous term in this fractional relaxation equation of motion. Here relaxation time is raised to the power $\alpha > 0$ in order to maintain the correct dimensionality.

Equations of the form (22) are mathematically well defined, and strategies for solving such equations have been developed by a number of investigators, particularly in the book by Miller and Ross[22] that is devoted almost exclusively to solving such equations when the index is rational. Here we make no such restriction and consider the Laplace transform of (22) to obtain

$$\widetilde{\Phi}(s) = \frac{\Phi(0)}{s} \frac{s^{\alpha}}{\lambda^{\alpha} + s^{\alpha}} \tag{23}$$

whose inverse Laplace transform is the solution to the fractional differential equation. Inverting Laplace transforms such as (23) is non-trivial and an excellent technique that overcomes many of the technical difficulties, implemented by Nonnenmacher and Metzler[23], involve the use of Fox functions[13]. The solution to the fractional relaxation equation is given by the series expansion for the standard Mittag-Leffler function

$$\Phi(t) = \Phi(0) E_{\alpha} \left(-(\lambda t)^{\alpha} \right) = \Phi(0) \sum_{k=0}^{\infty} \frac{(-1)^k}{\Gamma(1 + k\alpha)} (\lambda t)^{k\alpha} \tag{24}$$

which in the limit $\alpha \to 1$ yields the exponential function

$$\lim_{\alpha \to 1} E_{\alpha} \left(-(\lambda t)^{\alpha} \right) = e^{-\lambda t}$$

as it should, since under this condition (22) reduces to the ordinary relaxation rate equation (21).

The Mittag-Leffler function has interesting properties in both the short-time and the long-time limits. In the short-time limit it yields the Kohlrausch-Williams-Watts Law for stress relaxation in rheology given by

$$\lim_{t \to 0} E_\alpha \left(-(\lambda t)^\alpha \right) = e^{-(\lambda t)^\alpha} \tag{25}$$

also known as the stretched exponential. In the long-time limit it yields the inverse power law, known as the Nutting Law,

$$\lim_{t \to \infty} E_\alpha \left(-(\lambda t)^\alpha \right) = (\lambda t)^{-\alpha} . \tag{26}$$

Figure 1 displays the Mittag-Leffler function as well as the two asymptotes, the dashed curve being the stretched exponential and the dotted curve the inverse power law. What is apparent from this discussion is the long-time memory associated with the fractional relaxation process is inverse power law rather than the exponential of ordinary relaxation. It is apparent that the Mittag-Leffler function smoothly joins these two empirically determined asymptotic distributions.
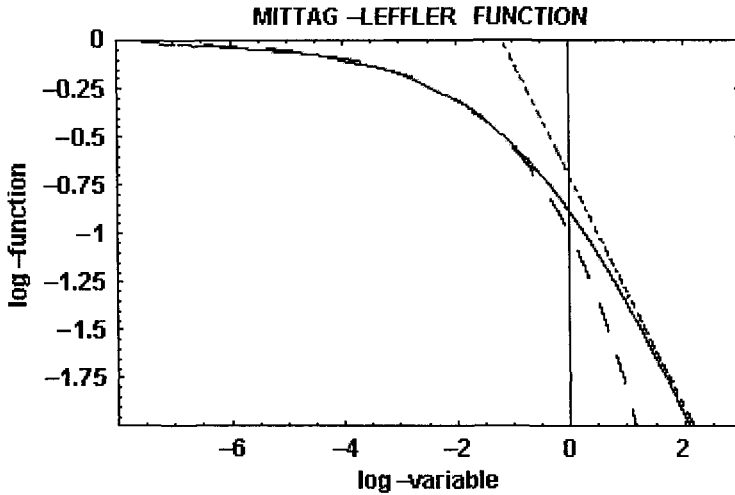
Figure1: The solid curve is the Mittag-Leffler function, the solution to the fractional relaxation equation. The dashed curve is the stretched exponential (Kohlrausch-Williams-Watts Law) and the dotted curve is the inverse power law (Nutting Law).

We can now generalize the fractional differential equation to include a random force $\xi(t)$ and in this way obtain a fractional Langevin equation

$$_0 D_t^{(\alpha)}[\Phi(t)] + \lambda^\alpha \Phi(t) = \frac{t^{-\alpha}\Phi(0)}{\Gamma(1-\alpha)} + \xi(t). \tag{27}$$

The solution to this equation is obtained using Laplace transforms as done previously

$$\tilde{\Phi}(s) = \frac{\Phi(0)s^{\alpha-1}}{\lambda^\alpha + s^\alpha} + \frac{\tilde{\xi}(s)}{\lambda^\alpha + s^\alpha}. \tag{28}$$

Note the difference in the s-dependence of the two coefficients of the right-hand side of (28). The inverse Laplace transform of the first term yields the Mittag-Leffler function as found in the homogeneous case above. The inverse Laplace transform of the second term is the convolution of the random force and a stationary kernel. The kernel is given by the series

$$E_{\alpha,\beta}(z) \equiv \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)}, \alpha > 0, \beta > 0. \tag{29}$$

which is the generalized Mittag-Leffler function. The function defined by (29) reduces to the usual Mittag-Leffler function when $\beta = 1$, so that both the homogeneous and inhomogeneous terms in the solution to the fractional Langevin equation can be expressed in terms of these series.

Note that taking the average value of (27) and observing that the average of the random force is zero, we obtain

$$_0 D_t^{(\alpha)}[\langle\Phi(t)\rangle] + \lambda^\alpha \langle\Phi(t)\rangle = \frac{t^{-\alpha}\Phi(0)}{\Gamma(1-\alpha)} \tag{30}$$

which is clearly of the form of the fractional stress relaxation equation. The average response of the system is determined by the Mittag-Leffler function, which is to say, the average has a long-time memory (inverse power law).

The explicit inverse of (27) yields the solution[13]

$$\Phi(t) = \Phi(0)E_\alpha\left(-(\lambda t)^\alpha\right) + \int_0^t (t-t')^{\alpha-1} E_{\alpha,\alpha}\left(-(\lambda t)^\alpha\right)\xi(t')dt' \tag{31}$$

In the case $\alpha = 1$, the Mittag-Leffler function becomes the exponential, so that the solution to the fractional Langevin equation reduces to that for an Ornstein-Uhlenbeck process

$$\Phi(t) = \Phi(0)e^{-\lambda t} + \int_0^t e^{-\lambda(t-t')}\xi(t')dt' \tag{32}$$

as it should. The analysis of the autocorrelation function of (31) can be quite daunting and so we do not pursue it further here, but refer the reader to the literature[13,24]. A somewhat simpler problem is the fractional Langevin equation without dissipation.

Consider the second moment of the solution to the Langevin equation when $\lambda = 0$ giving rise to

$$\langle[\Phi(t_1) - \Phi(0)][\Phi(t_2) - \Phi(0)]\rangle = \frac{1}{\Gamma(\alpha)^2}\int_0^{t_1}d\tau_1\int_0^{t_2}d\tau_2\frac{\langle\xi(\tau_1)\xi(\tau_2)\rangle}{(t_1-\tau_1)^{1-\alpha}(t_2-\tau_2)^{1-\alpha}}. \tag{33}$$

Here again we take the random force to have Gaussian statistics and to be delta correlated in time

$$\langle\xi(\tau_1)\xi(\tau_2)\rangle = C\delta(\tau_1 - \tau_2). \tag{34}$$

Inserting (34) into the expression for the autocorrelation function (33) and noting that the integral is symmetric in the times, the delta function restricts the integration to the lesser of the two times, so introducing the notation for the lesser time $t_<$ and greater time $t_>$ we obtain[13]

$$\langle[\Phi(t_>) - \Phi(0)][\Phi(t_<) - \Phi(0)]\rangle = \frac{2Ct_>^{\alpha-1}t_<^\alpha}{\Gamma(\alpha)^2}F\left(1;1-\alpha;1+\alpha:\frac{t_<}{t_>}\right) \tag{35}$$

in terms of the hypergeometric function. Note that although the statistics of the solution are Gaussian, they are also non-stationary, since the autocorrelation function depends on the lesser and the greater times separately and not on their difference.

Of course, we can also use the general expression (35) to write the second moment of the solution at time $t = t_< = t_>$,

$$\begin{aligned}\langle[\Phi(t) - \Phi(0)]^2\rangle &= \frac{2Ct^{2\alpha-1}}{\Gamma(\alpha)^2}F(1;1-\alpha;1+\alpha:1)\\ &= \frac{2Ct^{2\alpha-1}}{(2\alpha-1)\Gamma(\alpha)^2}\end{aligned} \tag{36}$$

where the second equality results from writing the hypergeometric function as the ratio of gamma functions. The time dependence of the second moment (36), if we make the identification $2H = 2\alpha - 1$, and the fractional index is less than one we have $1/2 \geq H > 0$. Consequently, the anomalous diffusion process described by the dissipation-free fractional Langevin equation is antipersistent.

This antipersistent behavior of the time series was observed by Peng et al.[25] for the differences in time intervals between heart beats. They interpreted their result, as did a number of subsequent investigators, in terms of random walks with $H < 1/2$. However, we can see from (36) that the fractional Langevin equation without dissipation is an equally good, or one might say an equivalent, description of the underlying dynamics. The scaling behavior alone cannot distinguish between these two models, what is needed is the complete statistical distribution and not just the time-dependence (scaling behavior) of a central moment.

3.1 Physical/physiological models

A physical Langevin equation is generally constructed from a Hamiltonian for a simple dynamical system coupled to the environment. The equations of motion for the coupled system are manipulated so as to eliminate the degrees of freedom of the environment from the dynamical description of the system. Only the initial state of the environment (heat bath) remains in the Langevin description, where the random nature of the driving force is inserted through the choice of distribution of the initial states of the heat bath. The simplest Langevin equation for a dynamical system open to the environment has the form

$$\frac{dX(t)}{dt} + \lambda X(t) = \xi(t) \tag{37}$$

where $\xi(t)$ is a random force, λ is a dissipation parameter and there exists a fluctuation-dissipation relation connecting the two[26]. Of course we can not completely interpret (37) until we specify the statistical properties of the fluctuations and for this we need to know the environment of the system. The random force is typically assumed to be a Wiener process, that is, to have Gaussian statistics and no memory.

When the system dynamics depends on what occurred earlier, that is, the environment has memory, (37) is no longer adequate and the Langevin equation must be modified. The generalized Langevin equation takes this memory into account through an integral term of the form

$$\frac{dX(t)}{dt} + \int_0^t K(t - t')X(t')dt' = \xi(t) \tag{38}$$

where the memory kernel replaces the dissipation parameter and the fluctuation-dissipation relation becomes generalized

$$K(\tau) = \langle \xi(t + \tau)\xi(t) \rangle. \tag{39}$$

Both Langevin equations (37) and (38) are monofractal if the fluctuations are monofractal, which is to say, the time series given by the trajectory $X(t)$ is a fractal random process if the random force is a fractal random process.

However, neither of these models is adequate for describing multifractal statistical processes as they stand. A number of investigators have recently developed multifractal random walk models to account for the multifractal character of various physiological phenomena and here we introduce a variant of those discussions based on the fractional calculus. The most recent generalization of the Langevin equation incorporates memory into the system's dynamics and has the simple form of (27) with the dissipation parameter set to zero

$$_0D_t^{(\mu)}[X(t)] - \frac{t^{-\mu}}{\Gamma(1-\mu)}X_0 = \xi(t).\tag{40}$$

The fractional Langevin equation (40) was obtained by Lutz[27] for a free particle coupled to a fractal heat bath, when the inertial terms is negligible. The formal solution to fractional Langevin equation (40) is

$$X(t) - X_0 = \frac{1}{\Gamma(1-\mu)}\int_0^t \frac{\xi(t')dt'}{(t-t')^{1-\mu}}$$

which can be expressed in terms of the integral kernel

$$\Delta X(t) \equiv X(t) - X_0 = \int_0^t K_\mu(t-t')\xi(t')dt'.\tag{41}$$

The form of relation (41) for multiplicative stochastic processes and its association with multifractals has been noted in the phenomenon of turbulent fluid flow[28], through a space, rather than time, integration kernel.

The random force term on the right-hand side of (41) is selected to be a zero-centered, Gaussian random variable and therefore to scale as[29]

$$\xi(\lambda t) = \lambda^H \xi(t)\tag{42}$$

where the Hurst exponent is in the range $0 < H \le 1$. In a similar way the kernel in (41) is easily shown to scale as

$$K_\mu(\lambda t) = \lambda^\mu K_\mu(t),\tag{43}$$

so that the solution to the fractional Langevin equation scales as

$$\Delta X(\lambda t) = \lambda^{H+\mu}\Delta X(t).\tag{44}$$

In order to make the solution to the fractional Langevin equation a multifractal we assume that the parameter µ is a random variable. To construct the traditional measures of multifractal stochastic processes we calculate the qth moment of the solution (44) by averaging over both the random force and the random parameter to obtain

$$
\begin{aligned}
\left\langle \left| \Delta X(\lambda t) \right|^q \right\rangle_{\xi,\mu} &= \lambda^{(q+1)H} \left\langle \lambda^{q\mu} \right\rangle_\mu \left\langle \left| \Delta X(t) \right|^q \right\rangle_\xi \\
&= \left\langle \left| \Delta X(t) \right|^q \right\rangle_\xi \lambda^{\varsigma(q)}
\end{aligned}
\tag{45}
$$

The scaling relation in (44) determines the qth order structure function exponent $\varsigma(q)$. Note that when ς (q) is linear in q the underlying process is monofractal, whereas, when it is nonlinear in q the process is multifractal, because we can relate the structure function to the mass exponent[30]

$$
\varsigma(q) = 2 - \tau(q).
\tag{46}
$$

Consequently we have that ς $(0) = H$ so that $\tau(0) = 2 - H$, as it should because of the well known relation between the fractal dimension and the global Hurst exponent $D_0 = 2 - H$.

To determine the structure function exponent we make an assumption about the statistics of the parameter µ. We can always write the µ-average as

$$
\left\langle \lambda^{q\mu} \right\rangle_\mu = \left\langle e^{qZ(\ln \lambda)} \right\rangle_z
\tag{47}
$$

where $Z(ln\lambda)$ is the random variable. In this way the expression on the right-hand side of (47) is the Laplace transform of the probability density. We assume the random variable is a α-stable Lévy process in which case the statistics of the multiplicative fluctuations are given by the distribution

$$
P(x,s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikz} e^{-bs|k|^\alpha} dk
\tag{48}
$$

with $0 < \alpha \le 2$. Inserting (48) into (47) and integrating over z yields the delta function $\delta(k+iq)$ which, integrating over k, results in

$$
\left\langle e^{qZ(\ln \lambda)} \right\rangle = e^{-b|q|^\alpha \ln \lambda} = \lambda^{-b|q|^\alpha}
\tag{49}
$$

so that comparing this result with (45) we obtain for the structure function exponent

$$
\varsigma(q) = (q+1)H - b|q|^\alpha.
\tag{50}
$$

Therefore the solution to the fractional Langevin equation corresponds to a monofractal process only in the case α = 1 and $q > 0$, otherwise the process is multifractal. We restrict the remaining discussion to positive moments.

Thus, we observe that when the memory kernel in the fractional Langevin equation is random, the solution consists of the product of two random quantities giving rise to a multifractal process. This is Feller's subordination process. We apply this approach to the stride interval time series depicted in Figure 2a.

Walking consists of a sequence of steps and the corresponding time series is made up of the time intervals for these steps. A stride interval, or step length, is the length of time from the one heel strike to the next heel strike of the same foot. It is the variability in the time series made from these intervals that is probed in our analysis of the stride interval time series. We refer to this variability as stride rate variability (SRV). We observe, using the SRV time series, that for the statistics of the multiplicative exponent given by Lévy statistics, the singularity spectrum as a function of the positive moments shown in Figure 2b.

The solid curve in Figure 2b is obtained from the analytic form of the singularity spectrum

$$f(q) = 2 - H - (\alpha - 1)bq^{\alpha}, \tag{51}$$

obtained by substituting the structure function exponent into the expression relating the structure function to the q-dependent mass exponent $f(q) = qh + \tau(q)$. It is clear from Figure 2 that the data for the singularity spectrum are well fit by a solution to the fractional Langevin equation with the parameters for the Lévy distribution given by $\alpha = 1.45$ and $b = 0.1$, obtained through a mean-square fit of (51) to the singularity spectrum determined by the SRV time series data.

(a)

Relaxed Walking
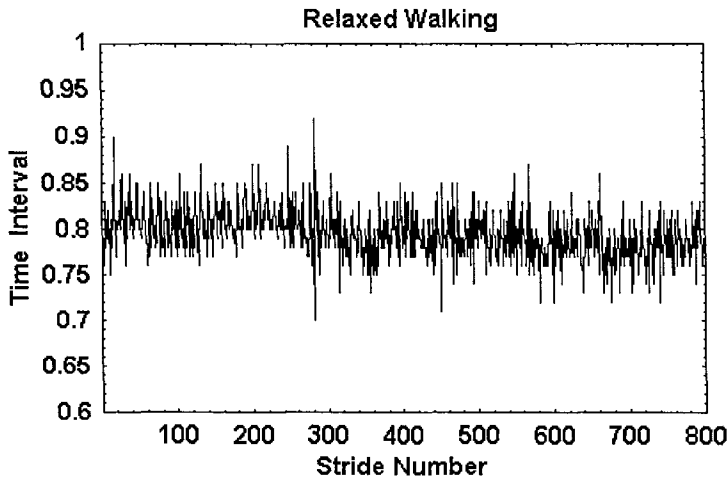
(b)
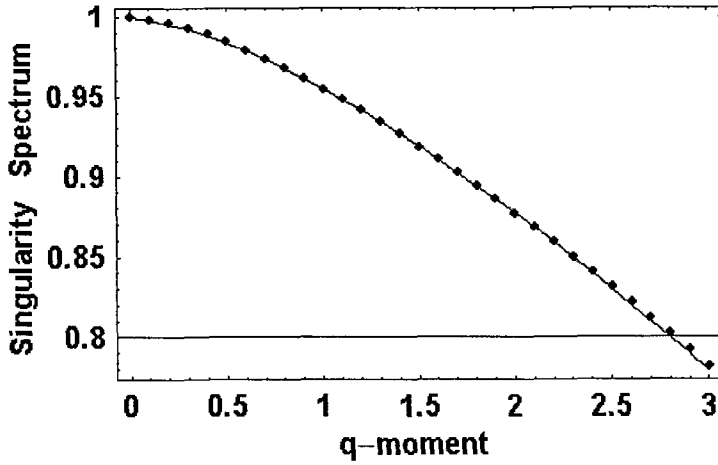
Figure 2: (a)The time interval between strides for the first 500 steps made by a typical walker in an experiment is depicted.(b) The singularity spectrum for $q > 0$ obtained through the numerical fit to the human gait data. The curve is the average over the ten data sets obtained in the experiment[31].

Of course, different physiologic processes generate different fractal time series, because the long-time memory of the underlying dynamical processes can be quite different. Physiological signals, such as cerebral blood flow (CBF), are typically generated by complex self-regulatory systems that handle inputs with a broad range of characteristics. Ivanov et al.[32] established that healthy human heartbeat intervals, rather than being fractal, exhibit multifractal properties and uncovered the loss of multifractality for a life-threatening condition of congestive heart failure. West et al.[33] similarly determined that CBF in healthy humans is also multifractal and this multifractality is severely narrowed for people who suffer from migraines.
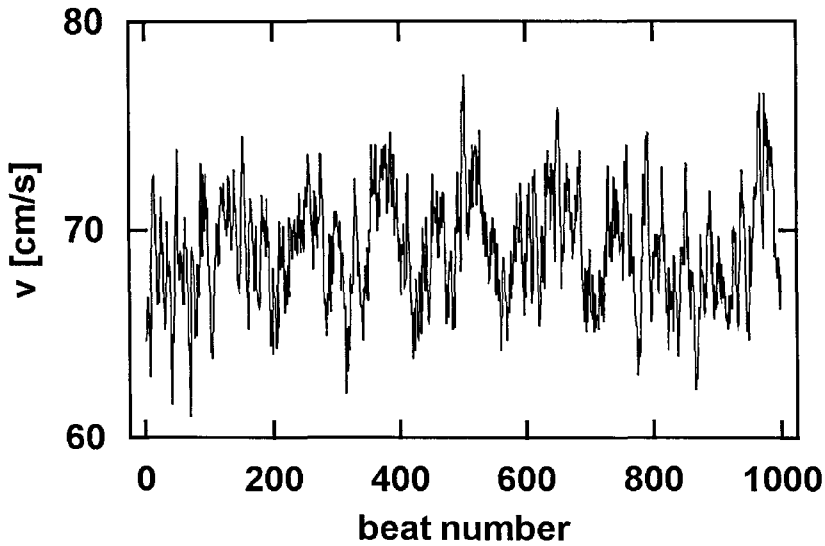
Figure 3: Middle cerebral artery flow velocity time series for a typical healthy subject[33].

Migraine headaches have been the bane of humanity for centuries, afflicting such notables as Caesar, Pascal, Kant, Beethoven, Chopin and Napoleon. However, its etiology and pathomechanism have to date not been satisfactorily explained. It was demonstrated[33] that the characteristics of CBF time series significantly differs between that of normal healthy individuals and migraineurs. Transcranial Doppler ultrasonography (TCD) enables high-resolution measurement of middle cerebral artery blood flow velocity. Like heart rate variability HRV, stride rate variability SRV and breathing rate variability BRV time series data, the time series of cerebral blood flow velocity consists of a sequence of waveforms. These waveforms are influenced by a complex feedback system involving a number of variables, such as arterial pressure, cerebral vascular resistance, plasma viscosity, arterial oxygen and carbon dioxide content, as well as other factors. Even though the TCD technique does not allow us to directly determine CBF values, it helps to clarify the nature and role of vascular abnormalities associated with migraine. In particular we present the multifractal properties of human middle cerebral artery flow velocity, an example of which is presented below in Figure 3.

In Figure 4 we compare the multifractal spectrum for middle cerebral artery blood flow velocity time series for a healthy group of five subjects and a group of eight migraineurs[33]. A significant change in the multifractal properties of the blood flow time series is apparent. Namely, the interval for the multifractal distribution on the local scaling exponent is greatly constricted. This is reflected in the small value of the width of the multifractal spectrum for the migraineurs 0.013, which is almost three times smaller than the width for the control group 0.038 for both migraineurs with and without aura.

The distributions are centered at 0.81, the same as that of the control group, so the average scaling behavior would appear to be the same.

(a)

(b)
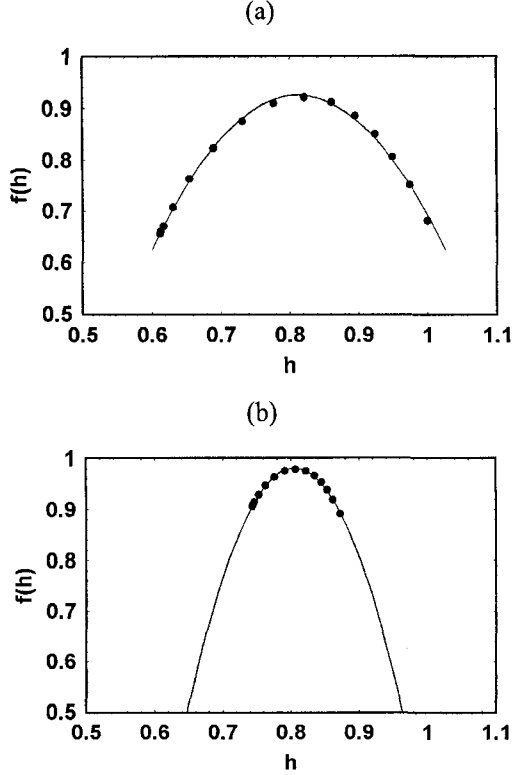
Figure 4: The average multifractal spectrum for middle cerebral blood flow time series is depicted by *f(h)*. (a) The spectrum is the average of ten time series measurements from five healthy subjects (filled circles). The solid curve is the best least-squares fit of the parameters to the predicted singularity spectrum. (b) The spectrum is the average of 14 time series measurements of eight migraineurs (filled circles). The solid curve is the best least-squares fit to the predicted spectrum using (53).

However, the contraction of the spectrum suggests that the underlying process has lost its flexibility. The biological advantage of multifractal processes is that they are highly adaptive, so that in this case the brain of a healthy individual adapts to the multifractality of the interbeat interval time series. We see that disease, in this case migraine, may be associated with the loss of complexity and consequently the loss of adaptability, thereby suppressing the normal multifractality of cerebral blood flow time series. Thus, the reduction in the width of the multifractal spectrum is the result of excessive dampening of the cerebral flow fluctuations and is the manifestation of the significant loss of adaptability and overall hyperexcitability of the underlying regulation system. West et al.[33] emphasize that hyperexcitability of the CBF control system seems to be

physiologically consistent with the reduced activation level of cortical neurons observed in some transcranial magnetic simulation and evoked potential studies.

Regulation of CBF is a complex dynamical process and remains relatively constant over a wide range of perfusion pressure via a variety of feedback control mechanisms, such as metabolic, myogenic, and neurally mediated changes in cerebrovascular impedance respond to changes in perfusion pressure. The contribution to the overall CBF regulation by different areas of the brain is modeled by the statistics of the fractional derivative parameter, which determines the multifractal nature of the time series. The source of the multifractality is over and above that produced by the cardiovascular system.

The multifractal nature of CBF time series is here modeled using a fractional Langevin model. We again implement the scaling properties of the random force and the memory kernel to obtain (51) as the scaling of the solution to the fractional Langevin equation. Here when we calculate the qth moment of the solution we assume Gaussian, rather than the more general Lévy, statistics. Consequently we obtain the quadratic function for the singularity spectrum

$$f(q) = 2 - H - bq^2 \tag{52}$$

which can be obtained from (51) by setting $\alpha = 2$. Another way to express (52) is

$$f(h) = f(H) - \frac{b}{4}(h - H)^2 \tag{53}$$

where we have used the fact that the fractal dimension is given by *2-H* which is the value of the function at *h = H*.

It seems that the changes in the cerebral autoregulation associated with migraine can strongly modify the multifractality of middle cerebral artery blood flow. The constriction of the multifractal to monofractal behavior of the blood flow depends on the statistics of the fractional derivative index. As the distribution of this parameter narrows down to a delta function, the nonlocal influence of the mechanoreceptor constriction disappears. On the other hand, the cerebral autoregulation does not modify the monofractal properties characterized by the single global Hurst exponent, presumably that produced by the cardiovascular system.

4. Summary, conclusions and speculations

Scaling suggests that the historical notion of disease, which has loss of regularity at its core, is inadequate for the treatment of dynamical diseases. Instead of loss of regularity, we identify the loss of variability with disease, so that a disease not only changes an average measure, such as stride rate, which it does in late stages, but is manifest in changes in stride rate variability at very early stages. Loss of variability implies a loss of

physiologic control and this loss of control is reflected in the change of fractal dimension, that is, in the scaling index of the corresponding time series[34].

The proper operation of physiologic processes is manifest through the scaling of appropriate time series. A measured function denoted by $X(t)$ is said to be homogeneous when the time axis being scaled by a constant γ yields the original function modified by an overall scale $X(\gamma t) = X(t)/\gamma^H$. This scaling behavior generalizes to time series when the measured function is stochastic and the scaling relation is interpreted in terms of the probability density function rather than the dynamic function itself.

We discussed how the fractional calculus could embody a number of the properties so prevalent in physiologic phenomena, not the least of which being that the evolution of a fractal processes can be described by a fractional differential equation. We showed that the evolution of fractal stochastic process can be described by a fractional Langevin equation in which a fractional differential equation is driven by a stochastic force. In particular, we demonstrated that the fractional calculus could provide a description of the dynamics of an anomalous diffusion process in which the long-time memory is not part of the stochastic driver, as it was in earlier models, but is actually part of the system's non-local dynamics through the fractional derivatives. The multifractal character of certain physiological time series, such as gait and cerebral blood flow, is described by fractional Langevin equations with random indices. The multifractal spectrum is shown to be related to the statistical properties of these random indices.

The well being of the body's system-of-systems is measured by the fractal scaling properties of the various dynamic subsystems and such scaling determines how well the overall complexity is maintained. Once the perspective that disease is the loss of complexity has been adopted, the strategies presently used in combating disease must be critically examined. Life support equipment is one such strategy, but the tradition of such life support is to supply blood at the average rate of the beating heart, to ventilate the lungs at their average rate and so on. So how does the new perspective regarding disease influence the traditional approach to healing the body?

Alan Mutch, of the University of Manitoba, argues that both blood flow and ventilation are delivered in a fractal manner in both space and time in a healthy body. However, during critical illness, conventional life support devices deliver respiratory gases by mechanical ventilation or blood by cardiopulmonary bypass pump in a monotonously periodic fashion. This periodic driving overrides the natural áperiodic operation of the body. Mutch speculates that these devices result in the loss of normal fractal transmission and consequently may do more damage the longer they are required[35]. Disease as the loss of complexity is consistent with the view that complex phenomena have a multiplicity of failure modes. These failure modes result in phenomena changing character, invariably becoming simpler with an accompanying inability to carryout their function. A cascade of failures is not so much a consequence of the initiating event, as it is the result of the state of the network when the event is initiated

A number of scientists[36] have demonstrated that the stability of hierarchal biological systems is a consequence of the interactions among the elements of the system.

Furthermore, there is an increase in stability resulting from the nesting of systems within systems – organelles into cells, cells into tissue, tissues into organs and so on up from the microscopic to the macroscopic. Each system level confers additional stability on the overall fractal structure. The fractal nature of the system suggests a basic variability in the way systems are coupled together. For example, the interaction between cardiac and respiratory cycles is not constant, but adapts to the physiologic challenges being experienced by the body.

We conclude with a number of observations: 1) The empirical evidence overwhelmingly supports the interpretation of the time series analysis that complex physiologic phenomena are described by fractal stochastic processes. Furthermore, the fractal nature of these time series is not constant in time but change with the vagaries of the interaction of the system with its environment and therefore these phenomena are multifractal. 2) The scaling index or fractal dimension marks the system's responsive and can be used as an indicator of the system's state of health. Since the fractal dimension is also a measure of the level of complexity, the change in dimension with disease suggests a new definition of disease as a loss of complexity, rather than the loss of regularity[34]. 3) The fractal dynamics of complex physiologic systems can be modeled using the fractional rather than the ordinary calculus because the changes in the fractal functions necessary to describe physiological complexity remain finite in the fractional formalism, but diverge in the traditional formalism[10].

References

1 J. Perrin, "Mouvement brownien et alit mol culaire", Annales de chimie et de physique VIII 18, 5-114: Translated by F. Soddy as *Brownian Movement and Molecular Reality*, Taylor and Francis, London.

2 E. Ott, *Chaos in Dynamical Systems*, Cambridge University Press (1993).

3 R.J. Abraham and C.D. Shaw, *Dynamics-The Geometry of Behavior, Part 1* (1982); *Part 2* (1983); *Part 3* (1985) and *Part 4* (1988), Aerial Press, Santa Cruz, CA.

4 B.J. West, *Physiology, Promiscuity and Prophecy at the Millennium: A Tale of Tails*, Studies of Nonlinear Phenomena in Life Science, Vol. 7, World Scientific, New Jersey (1999).

5 D. Ruelle, *Chaotic Evolution and Strange Attractors*, Cambridge University Press, Cambridge (1989).

6 M.F. Shlesinger, "Fractal time and 1/f noise in complex systems", *Ann. N.Y. Acad. Sci.* **504**, 214 (1987).

7 Yu. N. Rabotnov, *Elements of Hereditary Solid Mechanics*, MIR , Moscow (1980).

8 H.E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, Oxford (1979).

9 P. Meakin, *Fractals, scaling and growth far from equilibrium*, Cambridge Nonlinear Science Series 5, Cambridge University Press, Cambridge (1998).

10 M. Bologna, P. Grigolini and B.J. West, "Strange kinetics: conflict between density and trajectory description", *Chemical Physics* **284**, 115-128 (2002).

11 K.M. Kolwankar, "Studies of Fractal Structures and Processes using Methods of the Fractional Calculus", unpublished Thesis, University of Pune (1997).

12 This letter was translated by B. Mandelbrot and is contained in the Historical Sketches of his second book[3].

13 B.J. West, M. Bologna and P. Grigolini, *Physics of Fractal Operators*, Springer, New York (2003).

14 R. Hilfer, Editor, *Applications of Fractional Calculus in Physics*, World Scientific, Singapore (2000).

15 I. Sokolov, J. Klafter and A. Blumen, "Fractional Kinetics", *Physics Today*, Nov. 2002.

16 L.F. Richardson, "Atmospheric diffusion shown on a distance-neighbor graph", *Proc. Roy. Soc. London A* **110**, 709-737 (1926).

17 M.F. Shlesinger and B.J. West, "Complex fractal dimension of the bronchial tree", *Phys. Rev. Lett.* **67**, 2106-2109 (1991).

18 D. Sornette, "Discrete scale invariance and complex dimensions", *Phys. Rept.* **297**, 239-270 (1994).

19 A. Rocco and B.J. West, "Fractional calculus and the evolution of fractal phenomena", *Physica A* **265**, 535 (1999).

20 A. Rocco and B.J. West, "Fractional calculus and the evolution of fractal phenomena", *Physica A* **265**, 535 (1999).

21 T.F. Nonnenmacher and R. Metzler, "On the Riemann-Liouville fractional calculus and some recent applications", *Fractals* **3**, 557 (1995).

22 K.S. Miller and B. Ross, *An Introduction to the Fractional Calculus and Fractional Differential Equations*, John Wiley, New York (1993).

23 T.F. Nonnenmacher and R. Metzler, "On the Riemann-Liouville fractional calculus and some recent applications", *Fractals* **3**, 557 (1995).

24 V. Kobelev and E. Romanov, "Fractional Langevin equation to describe anomalous diffusion", *Prog. Theor. Phys. Supp.* **139**, 470-476 (2000).

25 C.K. Peng, J. Mistus, J.M. Hausdorff, S. Havlin, H.E. Stanley and A.L. Goldberger, "Long-range anticorrelations and non-Gaussian behavior of the heartbeat", Phys. Rev. Lett. **70**, 1343-46 (1993).

26 See for example, K. Lindenberg and B.J. West, *The Nonequilibrium Statistical Mechanics of Open and Closed Systems*, VCH, New York (1990).

27 E. Lutz, "Fractional Langevin Equation", *Phys. Rev. E* **64**, 051106 (2001).

28 D. Schertzer, S. Lovejoy, F. Schmitt, Y. Chigirinskays and D. Marsan, "Multifractal cascade dynamics and turbulent intermittency", *Fractals* **5**, 427 (1997).

29 J.B. Bassingthwaighte, L.S. Liebovitch and B.J. West, *Fractal Physiology*, Oxford University Press, New York (1994).

30 B. Rajagopalon and D.G. Tarboton, *Fractals* **1**, 6060 (1993).

31 B.J. West and M. Latka, "Fractional Langevin Model of Gait Variability", J. NeuroEng. Rehab.**2**:24 (2005).

32 P.C. Ivanov, L.A.N. Amaral, A.L. Goldberger, S. Havlin, M.G. Rosenblum, Z.R. Struzik, H.E. Stanley, "Multifractality in human heartbeat dynamics", *Nature* **399**, 461 (1999).

33 B.J. West, M. Latka, M.Galaubic-Latka and D. Latka, "Multifactility of cerebral blood flow", *Physica A* **318**, 453-460 (2003).

34 B.J. West, *Where medicine went wrong, rediscovering the path to complexity*, to be published (2005).

35 A. Mutch, "Health, 'Small-Worlds', Fractals and Complex Networks: An Emerging Field", *Med Sci Monit* **9**, MT55-MT59 (2003).

36 G.A. Chauvet, "Hierarchical functional organization of formal biological systems: a dynamical approach. I. The increase of complexity by self-association increases the domain of stability of a biological system", *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **339**(1290), 425-44 (1993).

This page is intentionally left blank

FRACTIONAL RELAXATION OF DISTRIBUTED ORDER

RUDOLF GORENFLO

*Department of Mathematics and Informatics, Free University Berlin,
Arnimallee 3, D-14195 Berlin, Germany
E-mail: gorenflo@mi.fu-berlin.de*

FRANCESCO MAINARDI*

*Department of Physics, University of Bologna and INFN
Via Irnerio 46, I-40126 Bologna, Italy
E-mail: mainardi@bo.infn.it*

The first-order differential equation of exponential relaxation can be generalized by re-placing the time-derivative by a fractional derivative of Caputo type, more generally by an integral over such fractional derivatives, the order of differentiation being the variable of integration. We give an outline of the theory, show the form of solutions for a few examples and investigate the asymptotics near zero and near infinity.

MSC 2000: 26A33, 26A48, 33E12, 44A10, 45J05, 45M05.

1. Introduction: statement of the problem and notations

The classical phenomenon of relaxation in its simplest form is known to be governed by a linear ordinary differential equation of order one, possibly non-homogeneous, that hereafter we recall with the corresponding solution. Denoting by $t \geq 0$ the time variable, $u = u(t)$ the field variable, and $q(t)$ a given continuous (or, more generally, integrable) function, that represents the possible forcing term, the *relaxation* differential equation (of non-homogeneous type) reads

$$u'(t) = -\lambda u(t) + q(t), \quad t \geq 0, \quad \lambda > 0, \tag{1}$$

where λ is a positive constant denoting the inverse of some characteristic time of the problem. The solution of (1), under the initial condition $u(0^+) = c_0$, is

$$u(t) = c_0 \, e^{-\lambda t} + \int_0^t q(t - \tau) \, e^{-\lambda \tau} \, d\tau \quad t \geq 0. \tag{2}$$

Such solution can be re-written as

$$u(t) = c_0 \, u_0(t) + (u_\delta * q)(t), \quad u_0(t) = e^{-\lambda t}, \quad u_\delta(t) = e^{-\lambda t} = -\frac{1}{\lambda} u_0'(t). \tag{3}$$

---

*Corresponding author

33

Here the function $u_0(t)$ represents the *fundamental solution*, namely the solution of the *homogeneous* equation which satisfies the initial condition $u_0(0^+) = 1$, and the function $u_\delta(t)$, with which the free term $q(t)$ appears convoluted, represents the so called *impulse-response solution*, namely the particular solution of the *inhomogeneous* equation with $c_0 \equiv 0$ and with $q(t) = \delta(t)$.

From the view-point of fractional calculus a natural generalization of equation (1) is obtained by replacing the ordinary derivative of order 1 with a fractional one of order $\beta \in (0, 1)$,

$$u^{(\beta)}(t) = -\lambda u(t) + q(t) \quad t \geq 0, \quad \lambda > 0, \quad 0 < \beta < 1, \tag{4}$$

subjected to the initial condition $u(0^+) = c_0$. In order to preserve the type of initial condition required in the classical phenomenon, the fractional derivative is meant in the Caputo sense, defined as follows according to three equivalent forms,

$$u^{(\beta)}(t) = \begin{cases} \dfrac{1}{\Gamma(1-\beta)} \displaystyle\int_0^t \dfrac{u^{(1)}(\tau)}{(t-\tau)^\beta}\, d\tau\,, \\ \dfrac{1}{\Gamma(1-\beta)} \dfrac{d}{dt}\displaystyle\int_0^t \dfrac{u(\tau)}{(t-\tau)^\beta}\, d\tau - \dfrac{t^{-\beta}}{\Gamma(1-\beta)}\, u(0^+)\,, \quad 0 < \beta < 1. \\ \dfrac{1}{\Gamma(1-\beta)} \dfrac{d}{dt}\displaystyle\int_0^t \dfrac{u(\tau) - u(0^+)}{(t-\tau)^\beta}\, d\tau\,, \end{cases} \tag{5}$$

In the limit $\beta = 1$ the Caputo derivative is known to reduce to the first derivative, but for $\beta = 0$ we find $u(t) - u(0^+)$.

For later use let us recall the following rule in the Laplace transform domain that can be used as the starting point to properly define the Caputo fractional derivative:

$$\mathcal{L}\left\{u^{(\beta)}(t); s\right\} = s^\beta\, \tilde{u}(s) - s^{\beta-1}\, u(0^+)\,, \quad 0 < \beta \leq 1, \quad \Re s > 0, \tag{6}$$

where $\tilde{u}(s)$ denotes the Laplace transform of $u(t)$.

We agree to refer to the equation (4) as the *simple fractional relaxation equation*. The reader interested to have more details on this topic may consult our review paper [1]. For treatises on fractional calculus we refer *e.g.* to [2,3,4].

The simple fractional relaxation equation can be generalized by using the notion *fractional derivative of distributed order*[a]. We thus consider the so-called *distributed order fractional relaxation equation* or *fractional relaxation equation of distributed order*,

$$\int_0^1 b(\beta)\, u^{(\beta)}(t)\, d\beta = -\lambda\, u(t) + q(t)\,, \quad t \geq 0, \quad \lambda > 0, \tag{7}$$

---

[a]We find a former idea of fractional derivative of distributed order in time in the 1969 book by Caputo [5], that was later developed by Caputo himself, see [6,7] and by Bagley & Torvik, see [8]. A basic framework for the numerical solution of distributed-order differential equations has been recently introduced by Diethelm & Ford [9].

subjected to the usual initial condition $u(0^+) = c_0$, where

$$b(\beta) \geq 0, \quad \text{and} \quad \int_0^1 b(\beta) \, d\beta = 1. \tag{8}$$

The normalization condition for the integral is here assumed for convenience but it is not mandatory. Clearly, some special conditions of regularity and behaviour near the boundaries will be required for the weight function $b(\beta)$, which, however is allowed to have δ components.

Being (4) and (7) linear non-homogeneous equations, though not of differential type like (1), it is easy to show the validity of the decomposition of their solutions into the two *basic solutions*, the fundamental solution $u_0(t)$ and the impulse response solution $u_\delta(t)$, that will no longer be of exponential type.

The purpose of this paper is to provide some preliminary results on the *basic solutions* of (7) both in general and in some special cases, results that are based on the application of the Laplace transformation. The plan of the paper is as follows.

In Section 2, by using the Laplace transform, we provide some relevant properties of the basic solutions $u_0(t)$ and $u_\delta(t)$ for the general case, namely properties that are independent from the choice of the weight function $b(\beta)$. In Section 3 we consider a few typical examples of weight functions $b(\beta)$, thereby concentrating our interest on the behaviour of $u_0(t)$ and $u_\delta(t)$ near zero and near infinity by using Tauberian theory. We note that finding explicit solution formulas rarely is possible. Finally, concluding remarks and directions for future work are given in Section 4.

2. Complete monotonicity of the basic solutions

Let us now apply the Laplace transform to Eq. (7) by using the rule (6) and introducing the relevant function

$$B(s) = \int_0^1 b(\beta) \, s^\beta \, d\beta. \tag{9}$$

We then get

$$B(s) \, \widetilde{u}(s) - \frac{B(s)}{s} \, c_0 = -\lambda \widetilde{u}(s) + \widetilde{q}(s),$$

namely

$$\widetilde{u}(s) = c_0 \, \widetilde{u}_0(s) + \widetilde{u}_\delta(s) \, \widetilde{q}(s), \tag{10}$$

where

$$\widetilde{u}_0(s) = \frac{B(s)/s}{\lambda + B(s)}, \quad \widetilde{u}_\delta(s) = \frac{1}{\lambda + B(s)}. \tag{11}$$

So the decomposition of the solution of the *fractional relaxation equation of distributed order* (7) into the two *basic solutions* is obtained in analogy with the the

classic relaxation equation (1), see the L.H.S of (3). The relation between these solutions, see the R.H.S of (3), is also verified. In fact, from the identity

$$\frac{\lambda}{\lambda + B(s)} = 1 - \frac{B(s)}{\lambda + B(s)},$$

and using (11) we have,

$$\widetilde{u}_\delta(s) = -\frac{1}{\lambda}\left(s\widetilde{u}_0(s) - 1\right), \tag{12}$$

hence, noting $u_0(0^+) = 1$,

$$u_\delta(t) = -\frac{1}{\lambda}u_0'(t), \tag{13}$$

like for the standard relaxation equation (1).

We are now going to show that the *basic solutions* $u_0(t)$ and $u_\delta(t)$, thanks to the conditions (8), are both *completely monotone*, so keeping the fundamental property of the corresponding basic solutions of the standard relaxation equation (1).

We recall that a function $f(t)$ defined for $t \geq 0$ is said to be *completely monotone* (CM) if it is a non-negative with infinitely many derivatives that alternate in sign. In other words

$$(-1)^n \frac{d^n}{dt^n} f(t) \geq 0, \quad n = 0, 1, 2, \ldots, \quad t \geq 0. \tag{14}$$

A necessary and sufficient condition is given by Bernstein's theorem, according to which a function $f(t)$ is CM in R^+ if and only if it is the restriction to the positive real semi-axis of the Laplace transform of a positive measure. For sake of simplicity we agree to represent $f(t)$ as

$$f(t) = \int_0^\infty e^{-rt} \rho(r)\, dr, \quad \rho(r) \geq 0, \tag{15}$$

where $\rho(r)$ is a non-negative (ordinary or generalized) function. Furthermore, a function $g(t)$ defined for $t \geq 0$ is said to be a *Bernstein function* if it is non-negative and its first derivative is a CM function. In other words

$$g(t) \geq 0, \quad (-1)^{n-1}\frac{d^n}{dt^n} g(t) \geq 0, \quad n = 1, 2, \ldots, \quad t \geq 0. \tag{16}$$

Thus the CM and Bernstein functions are special non-negative functions defined on the positive real semi-axis, the former being decreasing and convex, the latter increasing and concave. Relevant examples are the functions of exponential type, like $f(t) = e^{-t}$ and $g(t) = 1 - e^{-t}$, correspondingly. For more details on these classes of functions we refer *e.g.* to [10,11,12], where the reader can find a rigorous and exhaustive treatment. A more practical approach to this matter can be found in [13], from which we recall the following result, a sort of inverse of Bernstein's theorem: *The inverse Laplace transform of a CM function is non-negative.*

Let us now prove that the *basic solutions* $u_0(t)$ and $u_\delta(t)$ are CM functions. Let us first show that they are non-negative (as we naturally expect), as a consequence

of the above theorem. Looking into §9, Ch. II of [10] we are led to the following observations:

$B(s) = \int_0^1 b(\beta)\, s^\beta\, d\beta$ is Bernstein function because so is the function s^β for $s > 0$;

$B(s)/s = \int_0^1 b(\beta)\, s^{\beta-1}\, d\beta$ is CM function because so is the function $s^{\beta-1}$ for $s > 0$;

$\widetilde{u}_\delta(s) = 1/[\lambda + B(s)]$ is CM function, because it is reciprocal of a Bernstein function;
$\widetilde{u}_0(s) = B(s)/\{s[\lambda + B(s)]\}$ is CM function, because it is product of two CM functions.

As inverse Laplace transforms of CM functions, the *basic solutions* $u_0(t)$ and $u_\delta(t)$ are thus non-negative.

Are they completely monotone? To answer this question we invoke the Titchmarsh theorem on Laplace inversion. We observe the fact, that for $0 < \beta < 1$ the functions $s^{\beta-1}$ and s^β are regular in the main sheet of the Riemann surface (the complex plane cut along the negative real semi-axis), and that in particular for any $\mu > 0$ the functions $\mu + s^\beta$ never vanish there on the negative real semi-axis itself. Hence, if $b(\beta)$ does not contain δ-like contribution at $\beta = 0$ or $\beta = 1$, we have

$$u_0(t) = -\frac{1}{\pi} \int_0^\infty e^{-rt} \operatorname{Im}\left\{\widetilde{u}_0\left(re^{i\pi}\right)\right\} dr, \tag{17}$$

$$u_\delta(t) = -\frac{1}{\pi} \int_0^\infty e^{-rt} \operatorname{Im}\left\{\widetilde{u}_\delta\left(re^{i\pi}\right)\right\} dr, \tag{18}$$

and, according to the expressions of $\widetilde{u}_0(s)$ and $\widetilde{u}_\delta(s)$ in (11), we must look for the sign of

$$-\operatorname{Im}\left\{\frac{B(s)/s}{\lambda + B(s)}\right\}, \quad \text{and} \quad -\operatorname{Im}\left\{\frac{1}{\lambda + B(s)}\right\}$$

along the ray $s = re^{i\pi}$, $r > 0$.

For the second of these expressions we have

$$-\frac{1}{\lambda + B(s)} = -\frac{\lambda + \overline{B(s)}}{|\lambda + B(s)|^2},$$

so only the sign of $-\operatorname{Im}\left\{\overline{B(s)}\right\}$ along $s = re^{i\pi}$ is relevant. We have

$$-\operatorname{Im}\left\{\overline{\int_0^1 b(\beta)\, r^\beta\, e^{i\pi\beta}\, d\beta}\right\} = \int_0^1 b(\beta)\, r^\beta\, \sin(\pi\beta)\, d\beta > 0. \tag{19}$$

Hence, $u_\delta(t)$ is completely monotone, as Laplace transform (18) of a positive function.

Concerning the first expression, we observe

$$-\frac{B(s)}{s\,[\lambda + B(s)]} = -\frac{B(s)\left[\lambda + \overline{B(s)}\right]}{s\,|\lambda + B(s)|^2},$$

so only the signs of $-\mathrm{Im}\{B(s)/s\}$ and $-\mathrm{Im}\{1/s\}$ along $s = re^{i\pi}$ are relevant. Clearly, $-1/(re^{i\pi}) = 1/r > 0$. and from $B(s)/s = \int_0^1 b(\beta)s^{\beta-1}d\beta$ we conclude that

$$-\mathrm{Im}\left\{\int_0^1 b(\beta)\,r^{\beta-1}\,e^{i\pi(\beta-1)}\,d\beta\right\} = \int_0^1 b(\beta)\,r^{\beta-1}\sin(\pi\beta)\,d\beta > 0\,. \qquad (20)$$

Hence, $u_0(t)$ is completely monotone, as Laplace transform (17) of a positive function. Considering equations (13) and (14) we again see that also the function $u_\delta(t)$ is completely monotone.

3. Examples

Let us consider some typical choices for the weight function $b(\beta)$ that characterizes the relaxation equation of distributed order. Since finding explicit solution formulas may be possible only in a few cases we shall concentrate our interest on the asymptotic behaviour of $u_0(t)$ and $u_\delta(t)$ as $t \to 0$ and $t \to +\infty$ by using the Tauberian theory for Laplace transforms. According to this theory the asymptotic behaviour of a function $f(t)$ near $t = \infty$ and $t = 0$ is (formally) obtained form the asymptotic behaviour of its Laplace transform $\widetilde{f}(s)$ for $s \to 0^+$ and for $s \to +\infty$, respectively. For the application of this theory we limit ourselves to consider the behaviour of $u_0(t)$ since for $u_\delta(t)$ we may apply the differential relationship (12). Furthermore we shall use the asymptotic representations

$$\widetilde{u}_0(s) = \frac{B(s)/s}{\lambda + B(s)} \sim \begin{cases} \dfrac{1}{\lambda}\dfrac{B(s)}{s}\,, & s \to 0^+\,, \quad \text{being} \quad B(s) << \lambda\,, \\[2ex] \dfrac{1}{s}\left[1 - \dfrac{\lambda}{B(s)}\right], & s \to +\infty\,, \quad \text{being} \quad B(s) >> \lambda\,. \end{cases} \qquad (21)$$

3.1. The single-order fractional relaxation

The choice

$$b(\beta) = \delta(\beta - \beta_1)\,, \quad \text{with} \quad 0 < \beta_1 < 1 \qquad (22)$$

corresponding to the simple fractional relaxation equation has been exhaustively treated in our 1997 review paper [1], see also [14,15,16]. For this choice let us briefly recall the main results that are here obtained inverting the Laplace transforms (11) taking into account that now

$$B(s) = s^{\beta_1}\,, \qquad (23)$$

hence

$$\begin{cases} \widetilde{u}_0(s) = \dfrac{s^{\beta_1-1}}{\lambda + s^{\beta_1}} \;\Longleftrightarrow\; u_0(t) = E_{\beta_1}\left(-\lambda t^{\beta_1}\right) = E_{\beta_1,1}\left(-\lambda t^{\beta_1}\right), \\[2ex] \widetilde{u}_\delta(s) = \dfrac{1}{\lambda + s^{\beta_1}} \;\Longleftrightarrow\; u_\delta(t) = -\dfrac{1}{\lambda}\dfrac{d}{dt}E_{\beta_1}\left(-\lambda t^{\beta_1}\right) = t^{-(1-\beta_1)}E_{\beta_1,\beta_1}\left(-\lambda t^{\beta_1}\right), \end{cases} \qquad (24)$$

Here $E_\mu(z)$ and $E_{\mu,\nu}(z)$ (with $\mu > 0$, $\nu > 0$ and $z \in \mathbf{C}$) denote the Mittag-Leffler function, in one and two parameters respectively, on which the reader may inform himself by consulting *e.g.* the books [17,18,19] and the articles [20,21]. Now let us recall the general results relevant for us, that is, the Laplace transform pair, see Eq. (1.80) in [2],

$$t^{\nu-1} E_{\mu,\nu} \left(-\lambda t^\mu\right) \div \frac{s^{\mu-\nu}}{s^\mu + \lambda} \tag{25}$$

and $E_{\mu,\nu}\left(-\lambda t^\mu\right)$ is CM function if $0 < \mu \le 1$ and $\nu \ge \mu$. This fact can be proved as a consequence of the property according to which, if $0 < \mu \le 1$ and $\nu \ge \mu$, the function $E_{\mu,\nu}(-x)$ is CM for $x > 0$, see [22,23], or of the Laplace transform pair (25), see [24]. The asymptotic behaviours of the basic solutions can be obtained without detouring on their explicit expression in terms of the Mittag-Leffler functions but invoking the Tauberian theory. Restricting our attention to the fundamental solution, we obtain

$$\widetilde{u}_0(s) \sim \begin{cases} \dfrac{s^{\beta_1-1}}{\lambda}, & s \to 0^+, \\[2mm] \dfrac{1}{s}\left(1 - \lambda s^{-\beta_1}\right), & s \to +\infty, \end{cases} \tag{26}$$

so

$$u_0(t) \sim \begin{cases} \dfrac{1}{\lambda}\dfrac{t^{-\beta_1}}{\Gamma(1-\beta_1)}, & t \to +\infty, \\[2mm] 1 - \dfrac{\lambda t^{\beta_1}}{\Gamma(1+\beta_1)}, & t \to 0^+. \end{cases} \tag{27}$$

3.2. *The double-order fractional relaxation*

We now consider the choice

$$b(\beta) = b_1\delta(\beta - \beta_1) + b_2\delta(\beta - \beta_2), \quad \text{with} \quad 0 < \beta_1 < \beta_2 < 1, \tag{28}$$

where the constants b_1 and b_2 are both positive, possibly restricted to the normalization condition $b_1 + b_2 = 1$. In this case we have

$$B(s) = b_1 s^{\beta_1} + b_2 s^{\beta_2}. \tag{29}$$

The Tauberian theory implies for the fundamental solution as follows, taking into account that s^{β_2} is negligibly small in comparison to s^{β_1} for $s \to 0^+$ and, viceversa, s^{β_1} is negligibly small in comparison to s^{β_2} for $s \to +\infty$,

$$\widetilde{u}_0(s) \sim \begin{cases} \dfrac{b_1}{\lambda} s^{\beta_1-1}, & s \to 0^+, \\[2mm] \dfrac{1}{s}\left(1 - \dfrac{\lambda}{b_2}s^{-\beta_2}\right), & s \to +\infty, \end{cases} \tag{30}$$

so

$$u_0(t) \sim \begin{cases} \dfrac{b_1}{\lambda}\dfrac{t^{-\beta_1}}{\Gamma(1-\beta_1)}, & t \to +\infty, \\[2mm] 1 - \dfrac{\lambda t^{\beta_2}}{b_2\,\Gamma(1+\beta_1)}, & t \to 0^+. \end{cases} \tag{31}$$

3.3. *The uniformly distributed order fractional relaxation*

We now consider the choice

$$b(\beta) = 1. \tag{32}$$

For this case we have

$$B(s) = \int_0^1 s^\beta \, d\beta = \frac{s-1}{\log s}, \tag{33}$$

hence

$$\widetilde{u}_0(s) = \frac{1}{s} \frac{s-1}{\lambda \log s + s - 1} = \frac{1}{s} - \frac{1}{s} \frac{\lambda \log s}{\lambda \log s + s - 1}. \tag{34}$$

Then, invoking the Tauberian theory for *slow varying functions* [b], a topic adequately treated in the treatise on Probability by Feller [13], we have

$$\widetilde{u}_0(s) \sim \begin{cases} \dfrac{1}{\lambda \, s \log(1/s)}, & s \to 0^+, \\ \dfrac{1}{s} - \dfrac{\lambda \log s}{s^2}, & s \to +\infty, \end{cases} \tag{35}$$

so

$$u_0(t) \sim \begin{cases} \dfrac{1}{\lambda \log t}, & t \to +\infty, \\ 1 - \lambda t \log(1/t), & t \to 0^+. \end{cases} \tag{36}$$

4. Conclusions

We have studied in detail the distributed order fractional relaxation equation, the structure of its solutions, and in a few cases their asymptotic behaviour near zero and near infinity. The remarkable fact is that, in case of non-trivial distribution of fractional time derivatives, the smallest order determines the behaviour near infinity, but the largest order governs the behaviour near zero. The investigated equation can be considered as a transparent model for more general time fractional evolution equations in a Hilbert or Banach space where our relaxation coefficient λ is replaced by an operator A, positive in appropriate sense and acting in such space. As a relevant example we cite the equation of distributed order fractional diffusion, where in the linear case the individual Fourier modes exhibit fractional relaxation. We are planning to extend our investigation to the study of the concrete interpretation of general subordination formulas, see *e.g.* [25,26], in terms of integral transforms and special functions, in particular those of Mittag-Leffler type. For the related problem of fractional diffusion of distributed order we refer the reader to some recent papers including [27,28,29,30].

---

[b]**Definition:** We call a (measurable) positive function $a(y)$, defined in a right neighbourhood of zero, *slowly varying at zero* if $a(cy)/a(y) \to 1$ with $y \to 0$ for every $c > 0$. We call a (measurable) positive function $b(y)$, defined in a neighbourhood of infinity, *slowly varying at infinity* if $b(cy)/a(y) \to 1$ with $y \to \infty$ for every $c > 0$. Examples: $(\log y)^\gamma$ with $\gamma \in \mathbf{R}$ and $\exp(\log y / \log \log y)$.

References

1. R. Gorenflo and F. Mainardi, Fractional calculus: integral and differential equations of fractional order, in: A. Carpinteri and F. Mainardi (Editors), *Fractals and Fractional Calculus in Continuum Mechanics*, Springer Verlag, Wien, 1997, pp. 223–276. [Reprinted in `http://www.fracalmo.org`]

2. I. Podlubny, *Fractional Differential Equations*, Academic Press, San Diego, 1999.

3. S.G. Samko, A.A. Kilbas and O.I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications*, Gordon and Breach, New York, 1993. Translation from the Russian edition, Nauka i Tekhnika, Minsk, 1987.

4. B.J. West, M. Bologna and P. Grigolini, *Physics of Fractal Operators*, Springer Verlag, New York, 2003.

5. M. Caputo, *Elasticità e Dissipazione*, Zanichelli, Bologna, 1969 (in Italian).

6. M. Caputo, Mean fractional-order derivatives differential equations and filters, *Ann. Univ. Ferrara, Sez VII, Sc. Mat.* **41** (1995) 73-84.

7. M. Caputo, Distributed order differential equations modelling dielectric induction and diffusion. *Fractional Calculus and Applied Analysis* **4** (2001) 421-442.

8. R.L. Bagley and P.J. Torvik, On the existence of the order domain and the solution of distributed order equations, *Int. J. Appl. Math.* **2** (2000) 865-882, 965-987.

9. K. Diethelm and N.J. Ford, Numerical solution methods for distributed order differential equations, *Fractional Calculus and Applied Analysis* **4** (2001) 531-542.

10. C. Berg and G. Forst, *Potential Theory on Locally Compact Abelian Groups*, Springer, Berlin, 1975.

11. G. Gripenberg, A.-O. Londen and O. Staffans, *Volterra Integral and Functional Equations* Cambridge University Press, Cambridge, 1990.

12. K.S. Miller and S.G. Samko, Completely monotonic functions, *Integr. Transf. and Spec. Funct.* **12** (2001) 389-402.

13. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2, 2-nd Edition, Wiley, New York, 1971.

14. M. Caputo and F. Mainardi, Linear models of dissipation in anelastic solids, *Riv. Nuovo Cimento* (Ser. II) **1** (1971) 161–198.

15. F. Mainardi, Fractional relaxation-oscillation and fractional diffusion-wave phenomena, *Chaos, Solitons & Fractals* **7** (1996) 1461–1477.

16. F. Mainardi, Fractional calculus: some basic problems in continuum and statistical mechanics, in: A. Carpinteri and F. Mainardi (Editors), *Fractals and Fractional Calculus in Continuum Mechanics*, Springer Verlag, Wien and New-York, 1997, pp. 291–348. [Reprinted in `http://www.fracalmo.org`]

17. A. Erdélyi, W. Magnus, F. Oberhettinger and F.G. Tricomi, *Higher Transcendental Functions*, Bateman Project, McGraw-Hill, New York, 1955, Vol 3. [Ch. 18: Miscellaneous Functions, pp. 206-227]

18. M.M. Djrbashian, *Integral Transforms and Representations of Functions in the Complex Plane* Nauka, Moscow, 1966 (in Russian). [There is also the transliteration as Dzherbashian]

19. O.I. Marichev, *Handbook of Integral Transforms of Higher Transcendental Functions, Theory and Algorithmic Tables*, Chichester, Ellis Horwood, 1983.

20. F. Mainardi and R. Gorenflo, On Mittag-Leffler type functions in fractional evolution processes, *J. Comput. & Appl. Mathematics* **118** (2000) 283-299.

21. R. Gorenflo, J. Loutschko, Yu. Luchko Computation of the Mittag-Leffler function $E_{\alpha,\beta}(z)$ and its derivatives, *Fractional Calculus and Applied Analysis* **5** No 4 (2002) 491-518.

22. W.R. Schneider, Completely monotone generalized Mittag-Leffler functions, *Exposi-*

tiones Mathematicae **14** (1996) 3-16.

23. K.S. Miller and S.G. Samko, A note on the complete monotonicity of the generalized Mittag-Leffler function, *Real Anal. Exchange* **23** (1997) 753-755.

24. A. Hanyga, Realizable constitutive equations in linear viscoelasticity, in A. Le Mehauté, J. A. Tenreiro Machado, J.-C. Trigeassou and J. Sabatier (Editors), *Fractional Differentiation and its Applications*, U-Books, 2005, pp. 353-364. [Selected papers, First IFAC Workshop on Fractional Differentiation and its Applications (FDA'04), Bordeaux (France) 19-21 July 2004.]

25. E. Bazhlekova, The abstract Cauchy problem for the fractional evolution equation, *Fractional Calculus and Applied Analysis* **1** (1998) 255-270.

26. E. Bazhlekova, Subordination principle for fractional evolution equations, *Fractional Calculus and Applied Analysis* **3** (2000) 213-230.

27. A.V. Chechkin, R. Gorenflo and I.M. Sokolov, Retarding subdiffusion and accelerating superdiffusion governed by distributed-order fractional diffusion equations, *Phys. Rev. E* **66** (2002) 046129/1-6.

28. A. V. Chechkin, R. Gorenflo, I. M. Sokolov, V. Yu. Gonchar, Distributed order time fractional diffusion equation, *Fractional Calculus and Applied Analysis* **6** (2003) 259-279.

29. M. Naber, Distributed order fractional subdiffusion, *Fractals* **12** (2004) 23-32.

30. I.M. Sokolov, A.V. Chechkin and J. Klafter, Distributed-order fractional kinetics, *Acta Physica Polonica* **35** (2004) 1323-1341. [http://arXiv.org/abs/cond-mat/0401146]

FRACTIONAL TIME: DISHOMOGENOUS POISSON PROCESSES VS. HOMOGENEOUS NON-POISSON PROCESSES

P. ALLEGRINI, F. BARBI AND P. GRIGOLINI *

Dipartimento di Fisica, Università di Pisa, Largo Pontecorvo 3, 56127, Pisa, Italy
Email: allegrip@df.unipi.it; barbifrancesco@yahoo.it; grigo@df.unipi.it

P. PARADISI

ISAC-CNR, Sez. Lecce, Str. Prov. Lecce-Monteroni Km. 1.2, 73100 Lecce, Italy
Email: paradisi@isac.le.cnr.it

We describe a form of *modulation*, namely a dishomogeneous Poisson process whose event rate changes sporadically and randomly in time with a chosen prescription, so as to share many statistical properties with a corresponding non-Poisson renewal process. Using our prescription the correlation function and the waiting time distribution between events are the same. If we study a continuous-time random walk, where the walker has only two possible velocities, randomly established at the times of the events, we show that the two processes also share the same second moment. However, the modulated diffusion process undergoes a dynamical transition between superstatistics and a Lévy walk process, sharing the scaling properties of the renewal process only asymptotically. The *aging* experiment – based on the evaluation of the waiting time for the next event, *given* a certain time distance between another previous event and the beginning of the observation – seems to be the key experiment to discriminate between the two processes.

1. Time series with inverse-power-law waiting times

In this paper we numerically and theoretically discuss the statistical properties of a time series $\{\tau_i\}$ generated from within two different approaches to complexity, *renewal* and *modulation*, in such a way as to produce in both ways the same time distribution density $\psi(\tau)$ asymptotically decaying with an inverse power law. In our numerical treatment we use the simplest form:

$$\psi(\tau) = (\mu - 1)\frac{T^{\mu-1}}{(\tau + T)^{\mu}},\tag{1}$$

with $2 < \mu < 3$. This distribution is properly normalized, and the parameter T, making this normalization possible, gives information on the lapse of time necessary to reach the time asymptotic condition when $\psi(\tau)$ becomes identical to an inverse power law.

---

*also at IPFC-CNR via Moruzzi 1 56124 Pisa, Italy, and CNS-UNT, P.O.Box 311427 Denton, TX 76203-1427, U.S.A.

We use the sequence $\{\tau_i\}$ to distribute events on the time axis. The first event occurs at $t = \tau_1$, the second at time $t = \tau_1 + \tau_2$, and so on. The time intervals between two consecutive events are called *laminar regions*, the reason being that the dynamical model here under study to illustrate renewal theory is an idealization of the celebrated Manneville map [1], with the time interval between two events representing the fluid regular state. We keep calling these time intervals laminar regions even when working with modulation.

For the theoretical discussion of this paper, it is useful to turn the sequence of times $\{\tau_i\}$ into a diffusion generating fluctuation $\xi(t)$. Each time laminar region is assigned through a coin tossing procedure either the value W or the value $-W$ ($W > 0$), thus creating a dichotomous fluctuation $\xi(t)$, which is interpreted as a stochastic velocity. We then shall study the time evolution of the coordinate $x(t)$, obeying the dynamic prescription

$$\dot{x} = \xi. \tag{2}$$

The central issue of this study of the two kind of diffusion processes (renewal or modulation) has to do with scaling, namely, the property:

$$p(x,t) = \frac{1}{t^\delta} F\left(\frac{x}{t^\delta}\right), \tag{3}$$

which is expected to hold true in the time asymptotic limit. Complexity, namely the departure from ordinary statistical mechanics, is signaled by either $\delta \neq 0.5$ or $F(y)$ departing from the Gaussian form, or by both properties.

2. Modulation vs. renewal

The modulation approach to complexity, meant as the foundation of non-exponential waiting time distributions, is based on conventional Poisson processes whose rates sporadically changes in time (dishomogenous processes). For instance, a double-well potential under the influence of white noise yields the exponential distribution of the time of sojourn in the two wells [2]. In the case of a symmetric double-well potential we have

$$\psi(t) = \lambda \exp(-\lambda t). \tag{4}$$

The parameter λ is determined by the Arrhenius formula $\lambda = k \exp[-Q/k_B\theta]$. In the case when either the barrier intensity Q [2] or temperature θ [3] are slowly modulated, the resulting waiting time distribution becomes a superposition of infinitely many exponentials. It is known [4] that a superposition of infinitely many exponentially decaying functions can generate an inverse power law. If values of τ_i, with different labels i are selected from different exponential distribution, the resulting process is no doubt renewal. To make it become a form of modulation theory, we have to draw a large sequence of time values, with the index i moving from i_λ to $i = i_\lambda + N_d$, with $N_d >> 1$, from the same exponential distribution $\psi(\tau) = \lambda \exp(-\lambda\tau)$.

In recent times, the term superstatistics has been coined [5] to denote an approach to non-canonical distributions, of any form, not only the Nutting (Tsallis) form, as in the original work of Beck [6]. We note that Cohen points out explicitly [5] that the time scale to change from a canonical distribution to another must be much larger than the time scale of each canonical process. Thus, we can qualify superstatistics as a form of modulation. According to the modulation theory, we write the waiting time distribution $\psi(t)$ under the following form

$$\psi(\tau) = \int d\lambda \Pi(\lambda) \lambda \exp(-\lambda t), \tag{5}$$

where $\Pi(\lambda)$ is the Γ distribution of order $\mu - 1$ given by

$$\Pi(\lambda) = \frac{T^{\mu-1}}{\Gamma(\mu-1)} \lambda^{\mu-2} \exp(-\lambda T). \tag{6}$$

This formula was proposed by Beck [6] and used in a later work [7].

After generating the sequence $\{\lambda_j\}$ according to (6), we create the sequence $\{\tau_i\}$ with the following protocol. For any number λ_j, we generate a sequence $\{\tau_i^{(\lambda_j)}\}$ with a distribution density $\psi(\tau) = \lambda_j \exp(-\lambda_j \tau)$. To realize modulation, we create the sequence $\{\tau_i\} \equiv \{\tau_1^{(\lambda_1)}, \tau_2^{(\lambda_1)}, \ldots \tau_{N_d}^{(\lambda_1)}, \tau_1^{(\lambda_2)}, \tau_2^{(\lambda_2)}, \ldots \tau_{N_d}^{(\lambda_2)}, \ldots\}$. This means that after any extraction of λ we draw a fixed number N_d of waiting times. Notice that according to the arguments of Cohen [5] for modulation to be identified with superstatistics, it is necessary to make N_d very large, virtually infinite. This serves the purpose of allowing the system to reach the local form of canonical equilibrium. In the case of this paper the condition of canonical equilibrium is replaced by a form of ordinary diffusion with ordinary scaling. This is the diffusion process generated by Eq. (2).

Our prescription yields for $\psi(\tau)$ the non-exponential form of Eq. (1). We compare the generated modulation process with the corresponding renewal one, namely with the same $\psi(\tau)$. This latter process can be numerically obtained by setting $N_d = 1$, or, simply, by reshuffling the series $\{\tau_i\}$. Notice that the renewal process can also be obtained by directly generating $\psi(\tau)$, without any recourse to Poisson distributions. In other words, the renewal process yields a time series which is homogeneous and non-Poissonian. Since $\mu > 2$, the renewal process is ergodic, and the correlation function of the fluctuation $\xi(t)$, $\langle \xi(t_1)\xi(t_2) \rangle$, with $t_2 > t_1$, can be evaluated by adopting the time average,

$$\langle \xi(t_1)\xi(t_2) \rangle = \lim_{T \to \infty} \frac{\int_0^T dt \xi(t)\xi(t + t_2 - t_1)}{T}. \tag{7}$$

Using renewal theory [8] we find for the normalized correlation function $\Phi_\xi(t)$

$$\Phi_\xi(\tau) \equiv \frac{1}{W^2} \langle \xi(t)\xi(t + \tau) \rangle, \tag{8}$$

the following analytical expression

$$\Phi_\xi(\tau) = \frac{1}{\langle \tau \rangle} \int_\tau^\infty d\tau (t - \tau)\psi(\tau) \implies \frac{d^2}{dt^2} \Phi_\xi(\tau) = \frac{\psi(\tau)}{\langle \tau \rangle}. \tag{9}$$

On the other hand the distribution density $\Pi(\lambda)$ is an equilibrium distribution allowing us to evaluate the correlation function through the prescription of Eq. (7) even if we realize the inverse power law distribution of Eq. (1) through modulation rather than by means of the renewal prescription. We note that the time spent by the system with a given λ is inversely proportional to λ. Thus, the use of the prescription of Eq. (7) yields for the stationary correlation function, the expressions

$$\Phi_\xi(\tau) = \frac{\int d\lambda \frac{\Pi(\lambda)}{\lambda} \exp(-\lambda\tau)}{\int d\lambda \frac{\Pi(\lambda)}{\lambda}}. \tag{10}$$

By integrating Eq. (2), squaring $x(t)$, averaging on the Gibbs ensemble, and exploiting the stationary nature of the correlation function $\langle \xi(t_1)\xi(t_2)\rangle = \langle \xi(0)\xi(t_2 - t_1)\rangle$, we obtain that $\langle x^2(t)\rangle = 2W^2 \int_0^t dt' \int_0^{t'} dt'' \Phi_\xi(t'')$.

We conclude that the condition $2 < \mu < 3$ makes our prescription to modulation to generate the same superdiffusion of the corresponding renewal process. As a consequence, studying the second moment of the diffusing variable x, through the average over the distribution $p(x,t)$ of the diffusing process is not sufficient to discriminate between renewal and modulation.

3. Continuous-Time-Random-Walk approach to diffusion

We now address the problem of diffusion generated by modulation and use the continuous-time random walk to derive the asymptotic scaling properties of the process. The problem of diffusion with renewal has been the object of earlier work [9] to which we refer the interested reader to appreciate the difference between renewal and modulation on this specific issue. Here we limit ourselves to noticing that the *Lévy scaling*, namely $\delta = 1/(\mu-1)$ and F of Eq. (3) being a symmetric Lévy $(\mu-1)$-stable function, is a property of the central part of the spreading distribution, which is limited by ballistic peaks forcing the diffusion process to yield multi-scaling.

We now prove how scaling is driven by the *crucial events*, defined as the change of λ in our modulation prescription. In fact, each new drawing from $\Pi(\lambda)$ of a value λ sets the rule that N_d times τ_i have to be drawn from the same exponential waiting time distribution, $\psi_\lambda(\tau) = \lambda \exp(-\lambda\tau)$. This generates a subsequence that for large N_d is very extended, thereby producing a diffusion process with diffusion coefficient D_λ inversely proportional to λ. This stochastic diffusion rule lasts for a time that is equal to the sum of the N_d times. The concept of event implies unpredictability, a property shared by the random drawing of a given λ, and by the successive N_d random drawings of the times τ_i's from the distribution of $\psi_\lambda(\tau)$ as well. However, according to our definition only the drawing of λ is *crucial*.

We define as $P(x,t)$ the probability density of moving by a quantity x, either positive or negative, during time t, as an effect of the N_d drawings from the same Poisson distribution. Analogously, with the symbol $P^{(n)}(x,t)$ we denote the probability density that the particle moves by the quantity x, as a result of n crucial events, occurring in such a way as to ensure that the particle moves by the quantity

x exactly in a time t. Note that after the occurrence of these n crucial events, the particle is in general located at a distance x' from the origin different from x. With $n+1$ events the particle might overshoot that position. Thus the particle might make the remaining trip using the value of λ drawn at the last crucial event, occurred at a time $t' < t$. In other words, we calculate $p(x,t)$, $i.e.$, the probability density of finding the particle in position x at time t using the formula

$$p(x,t) = \sum_{n=1}^{\infty} \int_0^{\infty} dt' \int_{-\infty}^{+\infty} P^{(n)}(x',t')N(x-x',t-t') + N(x,t). \qquad (11)$$

Here the quantity $N(x,t)$ denotes the probability density of moving by a quantity x in a time t with no crucial event occurring. The Fourier-Laplace transform of $p(x,t)$, $\hat{p}(k,u)$, is given by the expression

$$\hat{p}(k,u) = \frac{1}{1 - \hat{P}(k,u)}\hat{N}(k,u). \qquad (12)$$

This expression is of crucial importance to determine the resulting scaling. To reach this important result we have to define first the function $\psi^{(i)}(x,t)$. This is the probability density of moving by the quantity x in time t, with i time drawings from the same Poisson distribution. This important function reads

$$\psi^{(i)}(x,t) = \int_0^{\infty} d\lambda \psi_\lambda^{(i)}(x,t)\Pi(\lambda), \qquad (13)$$

where $\psi_\lambda^{(i)}(x,t)$ denotes the probability density of moving by x in time t as a result of drawing i numbers from the Poisson distribution corresponding to the same value of the parameter λ. The functions $\psi_\lambda^{(i)}(x,t)$ obey the following recursion relation:

$$\psi_\lambda^{(i)}(x,t) = \int_0^t dt' \int_{-\infty}^{+\infty} dx' \psi_\lambda^{(i-1)}(x',t')\psi_\lambda^{(1)}(x-x',t-t'), \qquad (14)$$

with

$$\psi_\lambda^{(1)}(x,t) \equiv \lambda \exp(-\lambda t)\frac{1}{2}\left[\delta(x-Wt) + \delta(x+Wt)\right]. \qquad (15)$$

In conclusion, the physical meaning of the function $\psi^{(i)}(x,t)$ of (13) corresponds to adopting the following procedure. With probability density $\Pi(\lambda)$ we draw a given λ. Then, we draw i values of time intervals, and we move the particle by a quantity x with these i drawings.

With these prescriptions we express the quantity $P(x,t) \equiv P^{(1)}(x,t)$ as follows

$$P^{(1)}(x,t) = \psi^{(N_d)}(x,t). \qquad (16)$$

In other words, the transition by the quantity x in time t generated by a single crucial event is derived from Eq. (14) with $i = N_d$. Finally, for $N(x,t)$ we have

$$N(x,t) = \int_0^{+\infty} d\lambda \Pi(\lambda)N_\lambda(x,t), \qquad (17)$$

and

$$N_\lambda(x,t) = \sum_{i=1}^{N_d} \int_0^t dt' \int_{-\infty}^{+\infty} dx' \psi_\lambda^{(i-1)}(x',t') \Psi_\lambda(x-x', t-t'), \qquad (18)$$

where with $\psi_\lambda^{(0)}(x,t)$ we denote the initial condition $\delta(x)\delta(t)$. In (18) there are no crucial events involved because the maximum numbers of time drawings is $N_d - 1$, not enough to generate a new crucial event, if we start at time $t = 0$ with a crucial event. As a result of this partial number of drawings the particle reaches position x' in time t'. The remainder portion of space $x - x'$ in the remainder portion of time $t - t'$ is traveled with no further time drawing, according to the prescription

$$\Psi_\lambda(x,t) = \frac{[\delta(x+Wt) + \delta(x-Wt)]}{2} \int_t^{+\infty} dt' e^{-\lambda t'}. \qquad (19)$$

To establish the scaling produced by the modulation approach, we have to evaluate the Fourier-Laplace transform of $\psi^{(i)}(x,t)$, $\Psi_\lambda(x,t)$ and $P(x,t)$. Using the convolution theorem we have:

$$\hat{\psi}^{(i)}(k,u) = \int_0^{+\infty} d\lambda \Pi(\lambda) \left[\frac{\lambda(u+\lambda)}{(u+\lambda)^2 + k^2 W^2} \right]^i \qquad (20)$$

and

$$\hat{\Psi}_\lambda(k,u) = \frac{(u+\lambda)}{(u+\lambda)^2 + k^2 W^2}. \qquad (21)$$

From Eqs. (20) and (16) we derive immediately

$$\hat{P}(k,u) = \int_0^{+\infty} d\lambda \Pi(\lambda) \left[\frac{\lambda(u+\lambda)}{(u+\lambda)^2 + k^2 W^2} \right]^{N_d}. \qquad (22)$$

The evaluation of the asymptotic properties, $u \to 0$ and $k \to 0$, is essential to assess the system scaling. If we make the assumption of sending u and k to 0 first, and then integrating over λ, we get $\hat{p}(k,u) \approx [u + k^2 W^2 \langle \lambda^{-1} \rangle_\lambda]^{-1}$ (implying $\delta = 0.5$), where $\langle \cdot \rangle$ denotes an average over an ensemble of Poisson processes, each of them characterized by a fixed value of λ, selected from the distribution $\Pi(\lambda)/\lambda$ and then kept fixed forever. We note, however, that $\mu < 3 \implies \langle \lambda^{-1} \rangle_\lambda = +\infty$, casting some doubts on the conclusion that $\delta = 0.5$ (superstatistics limit [6]).

We now prove that the scaling stemming from a single modulated trajectory is the same Lévy scaling that is found in renewal processes. In these processes $\hat{p}(k,u)$, according to the Lévy theory [10], is asymptotically $p(k,u) \approx 1 - au - b \cdot k^\alpha$, where a and b are contant numbers. Using (22) we get this asymptotic property provided that we show that $\hat{P}(k,0) \approx 1 - const \cdot k^\alpha$. Note that we have to prove that $\alpha = \mu - 1$. For this purpose we focus on Eq. (22), we set $u = 0$, and, for simplicity but without loss of generality, $T = 1$. Substituting $y = \lambda/k$ we obtain:

$$\hat{P}(k,0) = \frac{k^{\mu-1}}{\Gamma(\mu-1)} \int_0^\infty y^{\mu-2} e^{-Tky} \left(\frac{y^2}{y^2+1} \right)^{N_d} dy. \qquad (23)$$

At this point, we use the binomial expansion and the useful property:

$$\left(\frac{y^2}{y^2+1}\right)^{N_d} = \frac{y^{2N_d}}{\sum_{i=0}^{N_d}\binom{N_d}{i}y^{2i}} = 1 - \frac{\sum_{i=0}^{N_d-1}\binom{N_d}{i}y^{2i}}{\sum_{i=0}^{N_d}\binom{N_d}{i}y^{2i}}. \qquad (24)$$

Let us plug this expression into the integral, and divide it in two parts:

$$\frac{k^{\mu-1}}{\Gamma(\mu-1)}\left[\int_0^\infty y^{\mu-2}e^{-Tky}dy - \int_0^\infty \frac{\sum_{i=0}^{N_d-1}\binom{N_d}{i}y^{2i}}{\sum_{i=0}^{N_d}\binom{N_d}{i}y^{2i}}y^{\mu-2}e^{-Tky}dy\right]. \qquad (25)$$

The first term gives $\Gamma(\mu-1)/k^{(\mu-1)}$, and therefore the unity term in the zero-th order of $\hat{P}(k,0)$. Looking at the second term, note that the integrand does not diverge in zero, at the infinity either: the integral can be considered as a Laplace transform (in Tk) of a finite quantity. So, at this first order, we obtain a constant.

In conclusion, for the integral (23) we obtain the expression $1 - const \cdot k^{\mu-1}$, which is the same first order expansion as for the renewal process, thus yielding $\delta = 1/(\mu-1)$, a property expected to show up asymptotically, whereas the superstatistics scaling $\delta = 0.5$ corresponds to short time scales, where the scenario of a statistical ensemble of never-changing Poisson processes is very accurate.

4. Aging effects in renewal and modulation theories

We have dealt with a number of similarities between two approaches to inverse-power-law relaxation: a dishomogeneous Poisson process and a homogenous renewal one. This distinction is widely made when studying glass-forming liquids. Also the related phenomenon of *aging* has been known for a long time as a property of spin glasses and polymers [11]. The most popular manifestation of aging is given by two time correlation functions, depending on two times t_1 and $t_2 > t_1$ not only through their difference. The decay as a function of $t_2 - t_1$ becomes slower for larger t_1. Aging is thought to be determined by the fact that the system under study is out of equilibrium and that the regression to equilibrium is involves times larger than the observation time [12].

Here we have in mind *renewal aging*. For recent papers on this form of aging we refer the readers to the following few papers [13,14,15]. The renewal model studied herein yields renewal aging, whose immediate manifestation is revealed by the dependence of $\psi(t)$ on the observation time.

The quantity $\psi(t)dt$ is the probability that a laminar region beginning at $t = 0$ ends in the small interval $[t, t + dt]$. If we have at our disposal a Gibbs ensemble of sequences, all of which begin with the beginning of the first laminar region, located at $t = 0$, to derive experimentally $\psi(t)$ we have to observe the time at which the first laminar region of each sequence of the sample ends. This is equivalent to making *preparation* and *observation* at the same time. Let us imagine now that preparation is made at time $t = -t_a < 0$. The first measured waiting time is denoted by τ_1. The first waiting time, at variance with the observation of the successive waiting times, does not necessarily correspond to the total time duration of a laminar region.

The resulting histogram records time lengths that are generally smaller than those corresponding to preparing the system at time $t = 0$. Nevertheless, in the case when the waiting time distribution is exponential, both long and short time lengths are reduced by the same percent. Thus, turning the histogram into a normalized waiting distribution density has the effect of recovering the same exponential form. A renewal exponential process does not age. In the non-exponential case delaying the observation process has the effect producing a percent cut of the short-time laminar regions larger than that of the long-time laminar regions. As a consequence, with the normalization of the distribution, the weight of the short-time laminar regions is reduced and the weight of the long-time laminar regions is enhanced, thereby generating a slower decay of the survival probability $\Psi(t)$.

We numerically generated the $\Pi(\lambda)$ distribution and used this approach to generate trajectories, i.e., artificial sequences of random waiting times. We compared the results for the aging analysis of trajectories characterized by the same exponent μ of the power-law, but generated from the renewal process and from modulation processes with different N_d.

Given the sequence of waiting times and an aging time t_a, we compute the truncated waiting times, $i.e.$, the difference between each waiting time and t_a. When the waiting time is shorter than t_a, then we take the successive waiting times until their sum exceeds t_a. Then the truncated waiting time is defined as the difference between this sum and t_a. In this way, we obtain a sequence of truncated waiting times characterized by some *aged* probability density distribution $\psi_{t_a}(\tau)$. Aging is revealed if this distribution changes with t_a. As stated, the modulation process with $N_d = 1$ is expected to coincide with the prediction of renewal theory. In Figs. 1 and 2 we compare the renewal and the modulation process with different values of N_d. Rather than measuring the function ψ_{t_a}, we evaluate the *aged* survival probability Ψ_{t_a}, which is related to ψ_{t_a} by the relation:

$$\Psi_{t_a}(\tau) = \int_{\tau}^{\infty} \psi_{t_a}(\tau')d\tau' = 1 - \int_{0}^{\tau} \psi_{t_a}(\tau')d\tau'. \tag{26}$$

The left portion of Fig. 1 refers to the case $N_d = 1$. In this case, the choice of a given λ determines only one waiting time. As a consequence, any laminar region does not have anything in common with the earlier laminar regions, and modulation is indistinguishable from renewal. The right part of Fig. 1 shows that it is enough to set $N_d = 10$ to make the dotted lines significantly depart from the dashed lines. We see that the dotted lines indicate a faster decay than the dashed lines, which implies that with modulation the aging effect is already significantly reduced. With a further increase of N_d (see Figs. 2) this effect becomes more pronounced. We see, in fact, that the dotted lines tend to overlap with the brand new survival probability. We conclude that for $N_d \to \infty$ the aging effect is annihilated. We thus conclude that superstatistics, as a form of infinitely slow modulation, yields no aging.
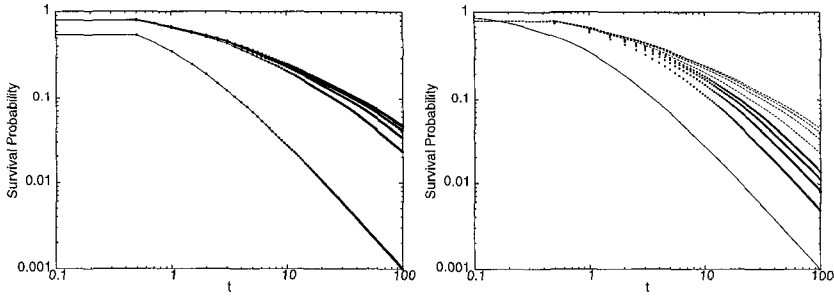
Figure 1. Left: Comparison between the function Ψ_{t_a} of a renewal process (continuous lines) and the Ψ_{t_a} of a modulation process with $N_d = 1$ (dots) at different values of $t_a = 0, 50, 100, 150, 200$ (from the lower to the upper curves). Right: Comparison between the function Ψ_{t_a} of a renewal process (dashed lines) and the function Ψ_{t_a} of a modulation process with $N_d = 10$ (dots), at different values of $t_a = 50, 100, 150, 200$ (from the lower to the upper curves). The lowest curve represents the brand new function $\Psi(t)$.
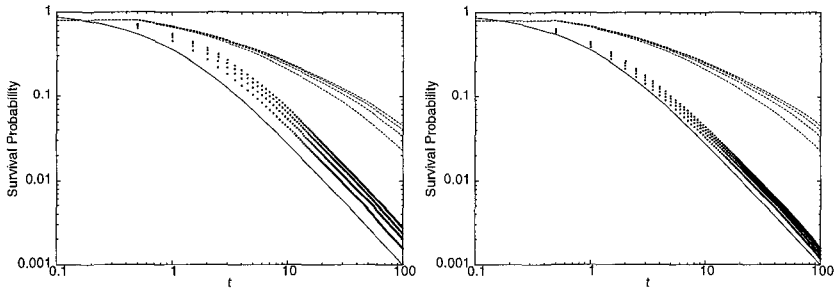
Figure 2. Left: The same as Fig. 1 (right), but with $N_d = 100$. Right: The same as Fig. 1, but with $N_d = 500$.

5. Concluding remarks

Let us discuss which are the main results emerging from this paper. First of all, we found that the waiting time distribution is not, by itself, an unambiguous indicator of complexity. This confirms the conclusion of earlier work [16] from within a new perspective made attractive by the modulation, or superstatistics, approach to complexity [5,6]. We have seen that the adoption of the aging experiment bypasses the limits of the conventional methods of analysis, since it establishes beyond any doubt the difference between the renewal and the modulation character of a sequence of times. This is a fact of some physical importance for the physics of blinking quantum dots [17]. The work of Brockmann et al. [18] has already established aging in the intermittent fluorescence of these new materials, and the more recent work of Ref. [19] confirms that this aging is due to the renewal character of the process. Thus, we conclude, in agreement with Ref. [19], that the dynamic process responsible for intermittent fluorescence must be built up on the basis of a renewal perspective.

Of course, the fact that superstatistics is ruled out as a proper perspective for the physics of blinking quantum dots does not mean, in any way, that there are no interesting complex processes in nature that might lay their roots on this attractive perspective. This remark is related to the second important result of this paper. Modulation, behind superstatitics, corresponds to a slow physical process whose realization does not rest on a unique procedure. However, it is a plausible that any realization of modulation might generate its own crucial events. These crucial events are embedded in a Poisson sea, exerting a camouflage action, which makes it extremely difficult, to reveal their existence. We have seen that these invisible events establish the anomalous scaling of the process. On the basis of the earlier work of Ref. [9] we think that this anomalous scaling refers to the central part of the probability density distribution, and that the resulting anomalous diffusion is multiscaling. There are good reasons to believe that with suitable developments along the lines of the work of Refs. [20,9], it will be possible to address successfully this challenging problem. This is left as a subject of future work.

References

1. P. Manneville, *J. Physique* **41**, 1235 (1980).
2. P. Allegrini, P. Grigolini, and A. Rocco, *Phys. Rev. E* **233**, 309 (1997).
3. M. Compiani, T. Fonseca, P. Grigolini, and R. Serra, *Chem. Phys. Lett.* **114**, 503 (1985).
4. M. F. Shlesinger and B. D. Hughes, *Physica A* **109**, 597 (1981).
5. E.G.D. Cohen, *Physica D* **193**, 35 (2004).
6. C. Beck, *Phys. Rev. Lett.* **87**, 180601 (2001).
7. M. Bologna, P. Grigolini, M. Pala, L. Palatella, *Chaos, Solitons and Fractals* **17**, 601 (2003).
8. T. Geisel, A. Zacherl, and G. Radons, *Phys. Rev. Lett.* **59**, 2503 (1987).
9. P. Allegrini, J. Bellazzini, G. Bramanti, M. Ignaccolo, P. Grigolini, J. Yang, Phys. Rev. E **66**, 015101(R) (2002).
10. E. W. Montroll and B. J. West, 'On an enriched collection of stochastic processes,' in *Studies in Statistical Mechanics Vol. vii: Fluctuation Phenomena*, ed. by E. W. Montroll and J. L. Lebowitz, North-Holland, Amsterdam (1979).
11. L. C. E. Struick, *Physical Aging in Amorphous Polymers and Other Materials* (Elsevier, Houston, 1978).
12. P. Calabrese and A. Gambassi, *J. Phys. A* **38** R 133 (2005).
13. G. Godréche and J. M. Luck, *J. Stat. Phys.* **104**, 489 (2001).
14. Eli Barkai *Phys. Rev. Lett.* **90**, 104101 (2003).
15. P. Allegrini, G. Aquino, P. Grigolini, L. Palatella, and A. Rosa, *Phys. Rev. E* **68**, 056123 (2003).
16. P. Allegrini, P. Grigolini, P. Hamilton, L. Palatella, and G. Raffaelli, *Phys. Rev. E* **65**, (2002).
17. M. Kuno, D.P. Fromm, H.F. Hamann, A. Gallagher, and D.J. Nesbitt, *J. Chem. Phys.* **115** (2001) 1028.
18. X. Brokmann, J.-P. Hermier, G. Messin, P. Desbiolles, J.-P. Bouchaud, M. Dahan, *Phys. Rev. Lett.* **90**, 120601 (2003).
19. S. Bianco, P. Grigolini, and P. Paradisi, submitted to *J. Chem. Phys.*
20. P. Grigolini, L. Palatella, G. Raffaelli, *Fractals* **9**, 439 (2001).

MARKOV MEMORY IN MULTIFRACTAL NATURAL PROCESSES

NIKITAS PAPASIMAKIS*, FOTINI PALLIKARI†

*Physics Department, Faculty of Physics, Department of Solid State Physics,
Panepistimiopolis Zografou, Athens 157 84, Greece*

Records of physical phenomena, such as turbulence, have been successfully modeled by random multiplicative processes. The present work expands such treatments by considering the effects of memory within the random multiplicative process and its consequences on the multifractal behavior of the measure. The measure-generating multiplicative cascade treated here involves first-order, two-state, Markov multipliers. When the two self-transition Markov probabilities, p_{ii}, $i = 1, 2$, are equal, yet different from 0.5, the average occurrence of the multipliers converges to 50% as in the memoryless case. Nevertheless, the Markov memory influences the spread of multipliers. The conservation of measure now relaxes to convergence towards a nontrivial and finite value and the shape of singularity spectrum depends to a great extent on the Markov probabilities. Application of the model to turbulence data indicates an underlying anti-persistent Markov process.

Keywords: multifractal, cascade, Markov, turbulence, persistence

1. Introduction

Complex physical processes, such as turbulence are known to be multifractal in nature [1-5]. Typical models of such processes are measures generated by multiplicative cascades, describing the constant redistribution of a measure from coarser to finer scales. The earliest and simplest multiplicative cascades involved independent fixed or random multipliers [6] on a tree-based structure and were not adequate to represent the underlying physical process, in contrast to more recent models which bring in an infinite number of random multipliers [7] and continuous scale invariance [8]. As it has been recognized that natural phenomena often exhibit short or long-range memory [9,10], there is good reason to assume that the redistribution of measure across different scales during the cascade should not take place through independent steps. However only recently the study of dependent cascades has come to focus in strictly mathematical sense [11] and on finite-state Markov generators [12] and turbulence [13-15].

---

*Email:npapasim@phys.uoa.gr
†Email:fpallik@phys.uoa.gr, fax number +30210-7276823

Markov sources are good models of a variety of physical processes, from the most familiar example of statistical mechanics, the Ising model, to the dynamics of Galton board processes [16-17]. Indeed the dynamics of the Galton board systems cover a wide spectrum of physical processes as for instance electronic solid-state diffusion [18], turbulent flow dynamics [19,20], optical interference [21], to name just a few. The analysis of multiplicative processes, where each multiplier is not just the random outcome of a probabilistic process, but of a Markovian memory-possessing source, can therefore find wide applications while providing useful insights to the underlying mechanisms of physical processes.

In this paper we combine the multifractal character of binomial multiplicative measures with the memory present in a first-order Markov process; the latter determines the transition probabilities in the generation of the two-state Markov multipliers. In particular, we investigate cases of persistence and anti-persistence [22] in the evolution of the Markov multipliers and the consequences on the multifractal characteristics of the binomial measure. In section 2 we review the construction of multifractal measures by multiplicative cascades and the resulting multifractal formalism. In section 3, the mathematical formalism of the two-state, first-order Markov process is reviewed for cases of persistence and anti-persistence. In section 4, we study the multifractal properties of multiplicative cascades driven by Markov processes considered in the previous section. Finally conclusions and applications are discussed in section 5.

2. Review of multifractal cascades

The original and most simple of multiplicative cascades, the binomial cascade, has been introduced as a model for the energy dissipation in fully developed turbulence [5]. To construct the binomial cascade one can start with an initial mass M_0 uniformly distributed over the interval $[0, 1]$ and two constants m_1 and m_2, the multipliers. At the first step of construction, the interval is divided into two sub-intervals, $[0, 1/2]$ and $(1/2, 1]$, and the initial mass M_0 is thus redistributed. The first sub-interval takes the mass $M_0 m_1$ and the second $M_0 m_2$. In the second step, each sub-interval is divided also into two new sub-intervals of equal length and its mass is also redistributed as was done in the first step. The cascade continues to the kth step where the initial mass is redistributed over a set of dyadic intervals of length 2^{-k}, the support of the measure. The mass contained in each sub-interval is thus predetermined by the current sequence of multipliers. In order for the measure to be conserved at each step, the multipliers must satisfy the simple relation

$$m_1 + m_2 = 1. \tag{1}$$

This type of multifractal measures, where the constructed cascade exactly prescribes the sequence of the multipliers for each sub-interval is thus a restricted one. The restriction can be alleviated and the multipliers randomized if they are allowed to occur at fixed probabilities, for example p and $1 - p$, respectively [23]. In this case

the exact conservation of the measure is relaxed to apply to the conservation of the measure on the average.

$$pm_1 + (1-p)m_2 = 1/2. \tag{2}$$

The total measure M plays now the role of a random variable (r.v.), Ω, with expected average, $E\{\Omega\} = 1$. In other words, in the limit where the number of construction steps tends to infinity, $k \to \infty$, the measure M is equal to the initial mass only on the average and for a large number of cascading sequences, k. Such cascades are called conservative or micro-canonical, in analogy to thermodynamics and we will refer to them as independent conservative cascades in contrast to the dependent Markov cases presented in section 4. As a further generalization to the case of fixed multipliers we consider two independent realizations of a random variable m [2]. The conservation of the measure on the average is satisfied in canonical multifractals for which

$$E\{m\} = 1/2. \tag{3}$$

In order to study the multifractal behavior of such measures, the qth moments $E\{[\mu_k]^q\}$ of the measure μ_k contained in dyadic intervals of length 2^{-k} are required. The measure μ_k equals the measure $\mu^{(k)}$ redistributed by the cascade to a random interval at the kth construction step multiplied by the r.v. Ω [a] ; $\mu_k = \mu^{(k)}\Omega$. Thus the qth moments can be written as

$$E\{[\mu_k]^q\} = E\{[\mu^{(k)}\Omega]^q\}. \tag{4}$$

Note that only in the case of restricted, or conservative, multifractals the r.v. Ω is always independent of $\mu^{(k)}$. Then a partition function $\tau(q)$ is defined through the relation

$$\tau(q) = -1 - \frac{\log_2[E\{[\mu^{(k)}\Omega]^q\}]}{k}. \tag{5}$$

And if the qth moments of Ω are positive, finite and independent of $\mu^{(k)}$ for every q, then at the limit $k \to \infty$, Ω can be omitted and equation (5) can be written as

$$\tau(q) = -1 - \frac{\log_2[E\{[\mu^{(k)}]^q\}]}{k}. \tag{6}$$

By appealing to the multifractal formalism found in [23-26] the scaling (Hölder) exponents, $h(q)$, are given by the derivative

$$h(q) = \frac{d\tau(q)}{dq} = -\frac{\log_2 e}{kE\{[\mu^{(k)}]^q\}}\frac{dE\{[\mu^{(k)}]^q\}}{dq}. \tag{7}$$

The spectrum function $f(q)$ is defined by the Legendre transform of $\tau(q)$ [25]

$$f(q) = qh(q) - \tau(q) = -q\frac{\log_2 e}{kE\{[\mu^{(k)}]^q\}}\frac{dE\{[\mu^{(k)}]^q\}}{dq} + \frac{\log_2[E\{[\mu^{(k)}]^q\}]}{k} + 1. \tag{8}$$

---

[a]Each dyadic interval at step k can be considered as a realization of the whole construction with initial mass $\mu^{(k)}$.

This review has considered for simplicity the dyadic grid, i.e. each interval is divided each time in 2 subintervals. Similarly, a generalization is possible for the case of an arbitrary b-adic grid, with an integer b as its base, i.e. each interval is divided each time in b subintervals.

3. The first-order, two-state Markov chain

The outcomes of random binary processes have a natural tendency to cluster just by chance [27]. In a Markov processes, however, this natural clustering can be further intensified or even weakened, depending on the type of memory present which connects the neighboring states in the evolution of the binary process. The transition probabilities in a two-state, first-order Markov process are shown in figure 1.
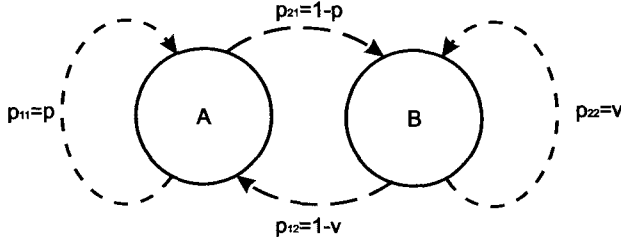
Figure 1.   State transition diagram of a first-order, two-state Markov process as a model of generation of cascade multipliers.

Once the outcome of a measurement finds the system to be at state A there is a probability p for the next outcome to also bring it to state A and $1 - p$ probability to bring it to state B. Once the outcome of a measurement finds the system to be at state B there is a probability ν for the next outcome to also yield state B and $1 - \nu$ probability to yield state A. If the absolute initial probabilities to find the system at either state A and B are $p_1^{(0)}$ and $p_2^{(0)}$, respectively, where $p_1^{(0)} = p_1$, then after n measurements, the probability to find the system at state A is [17]

$$p_n = a^{n-1}[p_1 - u] + u, \qquad (9)$$

where $u = \frac{1-\nu}{2-(p+\nu)}$, $a = p + \nu - 1$. The average expected occurrence of state A in the chain after n steps, p_n, would be written as

$$E\{p_n\} = u + \frac{p_1 - u}{n} \cdot \frac{1 - a^n}{1 - a} \underset{p=\nu}{=} 0.5 + \frac{p_1 - 0.5}{2n} \cdot \frac{1 - (2p - 1)^n}{1 - p}, \qquad (10)$$

where $0 < p < 1$ and $0 < \nu < 1$. From (9), (10) we conclude that the average expected occurrence of state A in the Markov chain converges to u which is 50% if $p = \nu > 0.5$ or equal to 50% for any number of steps, n, if the initial probability

to find the system at either states p_1 is 50%. The standard deviation of the mean value $E\{p_n\}$ is

$$\sigma = \sqrt{\frac{u(1-u)}{n}} \sqrt{\frac{1+a}{1-a}} \underset{p=\nu}{=} \frac{0.5}{\sqrt{n}} \sqrt{\frac{p}{1-p}}. \tag{11}$$

If $p = \nu = 0.5$ the Markov process is reduced to the Bernoulli, memoryless case. The presence of memory in the generation of states will make the scatter plot, $n = f(E\{p_n\})$, either to be broadened ($p > 0.5$, a case of persistence) or narrowed ($p < 0.5$, a case of anti-persistence) with respect to the random binomial sampling variation. We can now introduce the Markov source into the multiplicative cascades.

4. Introducing memory in the evolution of the multiplicative cascade

We consider again the process of the construction of the binomial measure as described above with positive multipliers m_1 and m_2. At each construction step the multipliers appear with probabilities described by a first-order Markov process and thus the probability of each multiplier depends on the preceding multiplier at the previous construction step. Following the notation introduced in section 1, we denote by p_{11} the probability that if at the present step of cascade the multiplier is m_1, in the next step the multiplier will also be m_1. Also we denote by p_{21} the probability that if at the present step of cascade the multiplier is m_1, in the next step the multiplier will be m_2, and so on. To exclude trivial cases, we will assume that none of the multipliers equals zero and that all probabilities are positive and smaller than one. After a large number of construction steps, m_1, m_2 appear in the cascade with probabilities $\frac{1-\nu}{2-p-\nu}$ and with $\frac{1-p}{2-p-\nu}$, respectively, equation (9). In the most interesting case where $p = \nu$, both probabilities equal 50%, but the cascade does not reduce to an independent conservative case, since the effect of memory could either introduce clustering of Markov multipliers, for $p > 0.5$, or their dispersion for $p < 0.5$. This naturally modifies the multifractal properties of the measure.

Our first concern is to examine the conditions under which the measure converges to a finite value. It is shown that the measure is conserved on the average at each step only if $p + \nu = 1$ or if at least one of the multipliers is zero. These cases are of no interest in the present work since the effect of memory is canceled. However the measure converges almost surely to a finite value [b] other than zero in the limit $k \to \infty$, when the following relation holds

$$4(1 - p - \nu)m_1 m_2 + 2(pm_1 + \nu m_2) = 1. \tag{12}$$

In the trivial case, where $p+\nu = 1$, relation (12) reduces to (2). In figure 2, examples of allowed pairs of multipliers are presented, both for the symmetrical case, (A), where $p = \nu$ and the asymmetrical case (B), where $p \neq \nu$. Note that the diagrams

---

[b]Note that this value is not necessary equal to the initial mass.
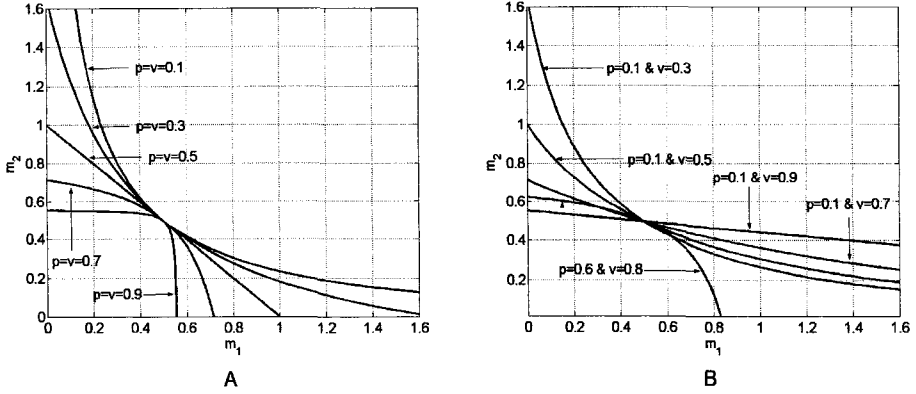
Figure 2. Examples of allowed pairs of multipliers for various values of Markov self-transition memories. (A) Symmetric case, $p = \nu$. (B) Asymmetric case, $p \neq \nu$.

are symmetric with respect to the bisector and that all lines intersect at the point $(1/2, 1/2)$, which corresponds to the trivial case where both multipliers are equal to $1/2$. Furthermore the multipliers cannot be both smaller than $1/2$ or greater than one, since in such cases the total measure would either be zero or infinity as $k \to \infty$. In the memoryless case, figure 2, the multipliers lie upon a straight line, instead.

Since the total measure Ω is also a random variable, its limiting probability distribution function would be required in order to fully describe the construction. However, if both multipliers are smaller than 1, then the moments $E\{\Omega_q\}$ are finite for every q [28]. We shall restrict our analysis to this case, while the more general case is reserved for a future work.

The next step is to study the qth moments of the measure $\mu^{(k)}$ assigned to a random dyadic interval at the kth construction stage of the cascade. By associating a binary r.v., which equals either $-\ln 2 \log_2[m_1]$ or $-\ln 2 \log_2[m_2]$, to the states of the Markov process, it can be shown by linear algebra that

$$E\{[\mu^k]^q\} = 2^{-k}\left[pm_1^q + \nu m_2^q + \sqrt{(pm_1^q + \nu m_2^q)^2 - 4(m_1 m_2)^q (p + \nu - 1)}\right]^k . \quad (13)$$

And since a Markov chain obeys a Large Deviation Principle [29], the multifractal formalism applies to our cascade. Furthermore the partition function is

$$\tau(q) = -\log_2\left[pm_1^q + \nu m_2^q + \sqrt{(pm_1^q + \nu m_2^q)^2 - 4(m_1 m_2)^q (p + \nu - 1)}\right], \quad (14)$$

where the factor containing the variable Ω has been omitted since its moments are finite. From Eq. (12) we have the anticipated property $\tau(1) = 0$ as a result of the conservation of the measure. Another special value of τ is $-\tau(0) = 1$, since $-\tau(0)$ equals the fractal dimension of the support of the measure, which in our case is 1.

If $m_2 < m_1 < 1$, then in the limit $q \to \infty$, $\tau(q)$ is given by

$$\tau(q \to \infty) = -1 - \log_2[p] - q\log_2[m_1], \tag{15}$$

and in the limit $q \to -\infty$ by

$$\tau(q \to -\infty) = -1 - \log_2[\nu] - q\log_2[m_2], \tag{16}$$

which implies that $\tau(q)$ is a linear function of q for very large positive and negative values of q. The Hölder exponents are given by

$$h(q) = -\frac{\log_2[m_1/m_2](pm_1^q - \nu m_2^q)}{2\sqrt{(pm_1^q + \nu m_2^q)^2 - 4(m_1 m_2)^q(p + \nu - 1)}} - \frac{log_2 m_1 m_2}{2}. \tag{17}$$

For $q = 0$ we obtain the most frequent scaling exponent

$$h(q = 0) = \frac{\log_2[m_1](\nu - 1) + \log_2[m_2](p - 1)}{2 - p - \nu}, \tag{18}$$

and in the case where $p = \nu$

$$h(q = 0) = -\frac{\log_2[m_1] + \log_2[m_2]}{2}. \tag{19}$$

By Eqs. (7), (15) and (16), the minimum value of the Hölder exponent h_{min} and the maximum h_{max} are

$$h_{min} = -\log_2 m_1 \quad \text{and} \quad h_{max} = -\log_2 m_2 \tag{20}$$

Thus since both multipliers are smaller than one and h is monotonous, $h(q) > 0$ for every q. The spectrum function $f(q)$ is

$$f(q) = -q\frac{\log_2[m_1/m_2](pm_1^q - \nu m_2^q)}{2\sqrt{(pm_1^q + \nu m_2^q)^2 - 4(m_1 m_2)^q(p + \nu - 1)}} - q\frac{log_2 m_1 m_2}{2}$$

$$+ \log_2\left[pm_1^q + \nu m_2^q + \sqrt{(pm_1^q + \nu m_2^q)^2 - 4(m_1 m_2)^q(p + \nu - 1)}\right] \tag{21}$$

And at the limits $q \to \pm\infty$

$$f(q \to \infty) = 1 + \log_2[p], \tag{22}$$

and

$$f(q \to -\infty) = 1 + \log_2[\nu]. \tag{23}$$

Since the shape of $f(h)$ is cap convex, we expect that negative dimensions will appear whenever at least one of the probabilities p, ν is smaller than $1/2$. On the other hand if p, ν are greater than $1/2$ then the minimum fractal dimension is positive. In figure 3 we present typical spectra for $p = \nu$ and $p \neq \nu$. It is obvious that the shape of the $f(h)$ curve depends on the values of p and ν. When $p = \nu$, $f(h)$ is symmetric, while if $p \neq \nu$, there is asymmetry between the right and left side. Furthermore, as the behavior of the cascade changes from anti-persistence to persistence, the spectrum becomes wider and more flat in order to obey equation (12).
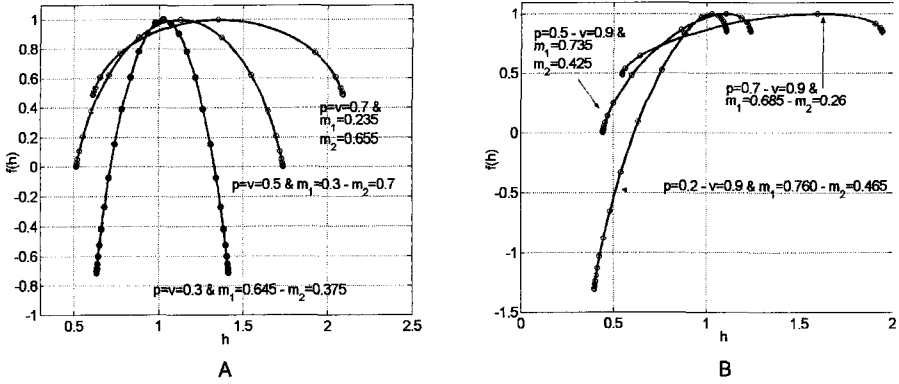
Figure 3. Multifractal spectra of Markov binomial cascades for $p = \nu$ (A) and $p \neq \nu$ (B). Solid circles: integer values of q .

5. Discussion

We have shown that by dropping the assumption of independence between the construction steps of a binomial conservative cascade, the multifractal properties of the resulting measure are significantly enriched. In particular, the shape of the multifractal spectrum depends only on the probabilities of the Markov chain and any asymmetry in the spectrum is a result of a non-symmetrical transition matrix. Persistence or anti-persistence in the underlying Markov process, leads to strictly positive values of the spectrum ($f(h) > 0$ for every q) or negative values ($f(h) < 0$ for some q), respectively. In contrast to the independent binomial cascade, these features may arise on both sides of the spectra at the same time. However, the most important innovation of the Markov cascade is the fact that the condition for convergence of the measure allows the multipliers to take a much wider range of values than in the independent case. Furthermore, while in the latter case, the measure is conserved on the average at each step, in the dependent case the conservation is relaxed to convergence, i.e. the measure converges to a finite value, which is not necessarily the initial mass.

Although these properties make the Markov cascade attractive for the modeling of a wide variety of natural processes, we will restrict our discussion to the case of energy dissipation in turbulence[c]. While the energy is conserved in 3 dimensions, this might not be the case across a one-dimensional cut. Additionally, the convergence of the measure ensures that the energy will not degenerate to zero or tend to infinity. As an example we will use the Markov cascade to model the results from an experiment performed in [30], where also negative values of $f(h)$

---

[c]Due to present shortage of space, a more in-depth study of relevance to experiments will be postponed for a future work.
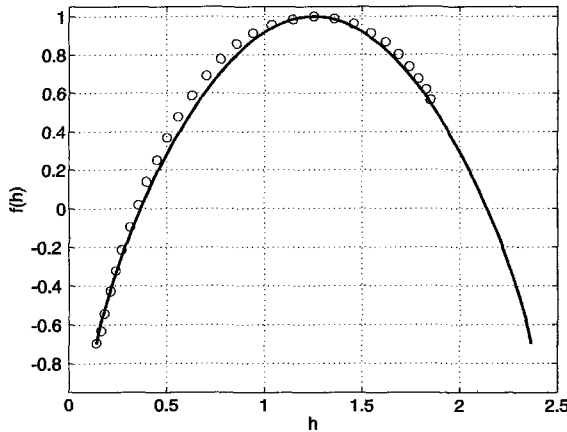
Figure 4. Multifractal spectrum. Open circles: experimental turbulence data of ref. [30]. Solid line: Markov cascade generated with self-transition probabilities $p = \nu = 0.308$, and multipliers $m_1 = 0.907$ and $m_2 = 0.194$.

were estimated using an improved box-counting algorithm [31]. In figure 4, the best fit to the spectrum (estimated from experimental data) is presented yielding $p = \nu = 0.308$, $m_1 = 0.907$ and $m_2 = 0.194$. The simulated and experimental spectra are in very good agreement and even though the latter is estimated only for $-2 < q < 5$, we nevertheless calculated the complete spectrum of the Markov cascade for $-\infty < q < \infty$. Such expansion is justified by the shape of the superimposed multifractal spectra of the experiments performed in [30]. The turbulence spectrum is symmetric and indicates an anti-persistent Markov cascade, where the multipliers alternate very rapidly. This behavior leads to very rare extremely large and small fluctuations and consequently to negative values of $f(h)$. For the estimated Markov parameters, the left part of relation (4.1) equals 0.95 which is very close to unity and consequently the measure will diverge slowly towards infinity as $k \to \infty$. Furthermore, the spectrum of the Markov cascade is somewhat narrower for small values of q (around the maximum of $f(h)$), which implies the need for a more complex model. We expect that a more general cascade driven by a multi-state, continuous Markov chain will improve the fit to experimental data and remove the constraint of discrete scale invariance, but this task is reserved for a future work.

Acknowledgments

62

References

1. R. Benzi, G. Paladin, G. Parisi, A. Vulpiani, *J. Phys. A Math. Gen.* **17**, 3521 (1984).
2. B. B. Mandelbrot, *Multifractals and 1/f Noise. Wild Self-Affinity in Physics*, Selecta volume, (Springer, New York, 1998).
3. H-O. Peitgen, H. Jurgens, D. Saupe, *Chaos and Fractals-New Frontiers of Science* (Springer, New York, 1992).
4. V. Yakhot, K. R. Sreenivasan, *nlin.CD/0406037* (2004)
5. C. Meneveau, K. R. Sreevinasan, *Phys. Rev. Lett.* **59**, 1424 (1987)
6. J. P. Kahane, J. Peyriere, *Adv. Math.* **22**, 131 (1976).
7. B. B. Mandelbrot , R. H. Riedi , *Adv. Appl. Math.* **16**, 132 (1995).
8. J. Barral, B. B. Mandelbrot, *Probab. Theory Relat. Fields* **124**, 409 (2002).
9. H. E. Hurst, *Transactions of the American Society of Civil Engineers* 116, 770 (1951).
10. F. Pallikari, *Chaos, Solitons and Fractals* **12**, 1499 (2001).
11. E. C. Waymire, S. C. Williams, *Bulletin Am. Math. Soc.* **31**, 216 (1994).
12. E. C. Waymire, S. C. Williams, *Markov cascades, Classical and Modern Branching Processes*, R. B. Athreya, P. Jagers eds, *Inst. Math. Appl.* IMA**84**, 305 (Springer-Verlag 1997).
13. A. Naert, R. Friedrich, J. Peinke, *Phys. Rev. E* **56**, 6719 (1997)
14. J. Cleve, M. Greiner, *Phys. Lett. A* **273**, 104 (2002)
15. I. Hosokawa , *Phys. Rev. E* **65**, (2002).
16. F. Galton, *Natural Inheritance.* (Macmillan, New York, 1894).
17. R. von Mises, *Mathematical Theory of Probability and Statistics*, (Academic Press, New York, 1964).
18. G. Hoover, B. Moran, C. Hoover, W. Ewans, *Phys. Lett. A* **133**, 114 (1988).
19. A. D. Chepelianskii, D. L. Shepelyansky, *Phys.Rev. Lett.* **87**, 034101, (2001).
20. A. Lue, H. Brenner, *Phys. Rev. E.* **47**, 3128 (1993).
21. P. Törma, *Phys. Rev. Lett.* **81**, 2185 (1998).
22. J. Feder, *Fractals*, (Plenum Press, New York 1988).
23. B. B. Mandelbrot, *Physica A* **163**, 306 (1990).
24. U. Frisch, G. Parisi, in *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics* (North-Holland, Amsterdam, 1985).
25. T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, B. I. Shraiman, *Phys. Rev. A* **33**, 1141 (1986).
26. R. H. Riedi, *J. Math. Anal. Appl.* **189**, 462 (1995).
27. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol 1, (Wiley, New York, 1950).
28. B. B. Mandelbrot, *J. Stat. Phys* **110**, 739 (2003).
29. R. S. Ellis, *Ann. Prob.* **12**, 1 (1984).
30. M. Alber, S. Luck, C. Renner, J. Peinke, *arxivnlin.CD/0007014* (2000)
31. M. Alber, J. Peinke, *Phys. Rev. E* **57**, 5487 (1998)

DESCRIPTION OF COMPLEX SYSTEMS IN TERMS OF SELF-ORGANIZATION PROCESSES OF PRIME INTEGER RELATIONS *

V. KOROTKIKH AND G. KOROTKIKH

Faculty of Informatics and Communication,
Central Queensland University,
Mackay, Queensland, 4740, Australia
E-mail: v.korotkikh@cqu.edu.au, g.korotkikh@cqu.edu.au

It is shown that complex systems can be described in terms of self-organization processes of prime integer relations. The prime integer relations are indivisible elements built from one level to the next by using the integers as the basic constituents. Prime integer relations specify nonlocal correlations between the parts of a complex system and can be represented as two-dimensional geometric patterns with fractal properties. Controlled by arithmetic the self-organization processes of prime integer relations describe complex systems by absolutely true information and do not require deeper principles for explanations. This gives the possibility to develop an irreducible theory of complex systems.

1. Introduction

Complex systems profoundly change human activities of the day. In order to understand and control the developments, it is becoming increasingly important to have confidence in the theory of complex systems. Ultimately, this requires clear explanations why the foundations of the theory of complex systems are valid in the first place. The ideal situation would be to have an irreducible theory of complex systems not requiring a deeper explanatory base in principle. But the question arises: where such an irreducible theory might come from at the time when even the concept of space-time is questioned [1] as a fundamental entity?

As an answer to the question, it is suggested that the concept of integers may take responsibility in the search for an irreducible theory of complex systems [2]. The aim of the paper is to present an irreducible description of complex systems in terms of self-organization processes of prime integer relations. A prime integer relation is an indivisible element in the sense that none of its components can be considered separately to result in a different prime integer relation. Controlled by arithmetic the self-organization processes characterize complex systems by absolutely true information and do not require further explanations. This gives the possibility to

---

develop an irreducible theory of complex systems.

2. Invariant Quantities of a Complex System and Correlations

We approach complex systems from a general perspective. In particular, we consider the dynamics of a complex system in order to understand correlations between the parts. For this purpose we describe a complex system by its certain quantities and are interested to find out how many of them remain invariant as the system evolves from one state to another [2], [3].
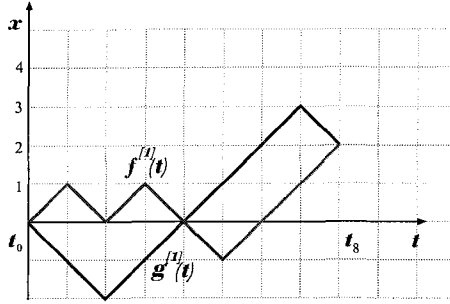
Figure 1. For states $x = +1 - 1 + 1 - 1 - 1 + 1 + 1 + 1$ and $x' = -1 - 1 + 1 + 1 + 1 + 1 + 1 - 1$ we have $f^{[1]}(t_8) = g^{[1]}(t_8)$, where $f = \rho_{m \varepsilon \delta}(x), g = \rho_{m \varepsilon \delta}(x')$ and $m = 0, \varepsilon = 1, \delta = 1$.

Let I be an integer alphabet and $I_N = \{x = x_1...x_N, x_i \in I, i = 1, ..., N\}$ be the set of sequences of length $N \geq 2$. We consider a complex system of N elementary parts with the state of part P_i specified by a space variable $x_i \in I, i = 1, ..., N$ and the state of the complex system itself by a sequence $x = x_1...x_N \in I_N$.
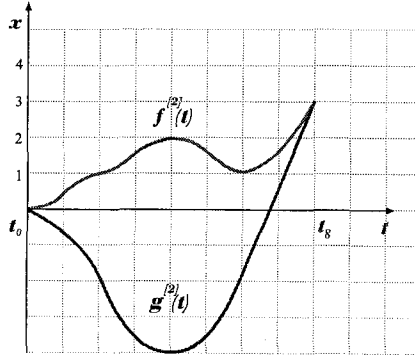
Figure 2. For the states x and x' we also have $f^{[2]}(t_8) = g^{[2]}(t_8)$. It is clear that $f^{[3]}(t_8) \neq g^{[3]}(t_8)$ and thus $C(x, x') = 2$.

We use piecewise constant functions for a geometric representation of the sequences. Let $\varepsilon > 0$ and $\delta > 0$ be length scales of a two-dimensional lattice. Let $\rho_{m\varepsilon\delta} : x \rightarrow f$ be a mapping that realizes the geometric representation of a sequence $x = x_1...x_N \in I_N$ by associating it with a function $f \in W_{\varepsilon\delta}[t_m, t_{m+N}]$, denoted as $f = \rho_{m\varepsilon\delta}(x)$, such that $f(t_m) = x_1\delta$, $f(t) = x_i\delta$, $t \in (t_{m+i-1}, t_{m+i}]$, $i = 1, ..., N$, $t_i = i\varepsilon$, $i = m, ..., m + N$ and whose integrals $f^{[k]}$ satisfy $f^{[k]}(t_m) = 0$, $k = 1, 2, ...$, where m is an integer. The sequence $x = x_1...x_N$ is called a code of the function f and denoted by $c(f)$.

We use the geometric representation to characterize a state $x = x_1...x_N \in I_N$ of a complex system in terms of the definite integrals

$$f^{[k]}(t_{m+N}) = \int_{t_m}^{t_{m+N}} f^{[k-1]}(t)dt, \ f^{[0]} = f, \ k = 1, 2, ... \tag{1}$$

of a function $f = \rho_{m\varepsilon\delta}(x) \in W_{\varepsilon\delta}[t_m, t_{m+N}]$ [2]. The definite integrals specify certain quantities of the complex system.

The integer code series [4] expresses the quantities (1) of a complex system in terms of the state $x = x_1...x_N \in I_N$ explicitly. Namely, it describes the definite integral

$$f^{[k]}(t_{m+N}) = \sum_{i=0}^{k-1} \alpha_{kmi}((m + N)^i x_1 + ... + (m + 1)^i x_N)\varepsilon^k \delta \tag{2}$$

of a function $f \in W_{\varepsilon\delta}([t_m, t_{m+N}])$ by using the code $c(f) = x_1...x_N$ of the function f, powers $(m + N)^i, ..., (m + 1)^i$ of integers $(m + N), ..., (m + 1)$ and combinatorial coefficients

$$\alpha_{kmi} = ((-1)^{k-i-1}(m + 1)^{k-i} + (-1)^{k-i}m^{k-i})/(k - i)!i!,$$

where $k \geq 1$ and $i = 0, ..., k - 1$ [4].

We are concerned how many quantities (1) remain invariant as a complex system moves from one state $x = x_1...x_N \in I_N$ to another $x' = x'_1...x'_N \in I_N$ [2]

$$f^{[k]}(t_{m+N}) = g^{[k]}(t_{m+N}), \quad k = 1, ..., C(x, x'), \tag{3}$$

$$f^{[C(x,x')+1]}(t_{m+N}) \neq g^{[C(x,x')+1]}(t_{m+N}), \tag{4}$$

where $f = \rho_{m\varepsilon\delta}(x), g = \rho_{m\varepsilon\delta}(x')$ (Figures 1 and 2). We consider the conservation of the quantities (3) as a consequence of correlations between the parts of the complex system. In particular, although the complex system moves from one state to another, the changes of the parts, nevertheless, are correlated to preserve the quantities. The more of the quantities are preserved the stronger the correlations between the changes of the parts.

3. The Conserving Correlations as Linear Equations

As a complex system evolves from one state $x = x_1...x_N \in I_N$ to another $x' = x'_1...x'_N \in I_N$, it is proved [2], by using (2), that the number of its $C(x, x') \geq 1$ quantities remain invariant, iff the space-time changes $\Delta x_i = x'_i - x_i, i = 1, ..., N$ of the elementary parts are correlated to satisfy $C(x, x')$ linear equations

$$(m + N)^k \Delta x_1 + ... + (m + 1)^k \Delta x_N = 0, \tag{5}$$

$k = 0, ..., C(x, x') - 1$ and an inequality

$$(m + N)^{C(x,x')} \Delta x_1 + ... + (m + 1)^{C(x,x')} \Delta x_N \neq 0. \tag{6}$$

Moreover, the number $C(x, x')$ of the conserved quantities of a complex system satisfies the condition $0 \leq C(x, x') < N$ [2]. The coefficients of the linear equations (5) become the entries of the matrix

$$\begin{pmatrix} (m + N)^0 & (m + N - 1)^0 & ... & (m + 1)^0 \\ (m + N)^1 & (m + N - 1)^1 & ... & (m + 1)^1 \\ . & ... & & . \\ (m + N)^{N-1} & (m + N - 1)^{N-1} & ... & (m + 1)^{N-1} \end{pmatrix}$$

with Vandermonde determinant, when the number of the equations is N. This fact is important in order to prove that $C(s, s') < N$.
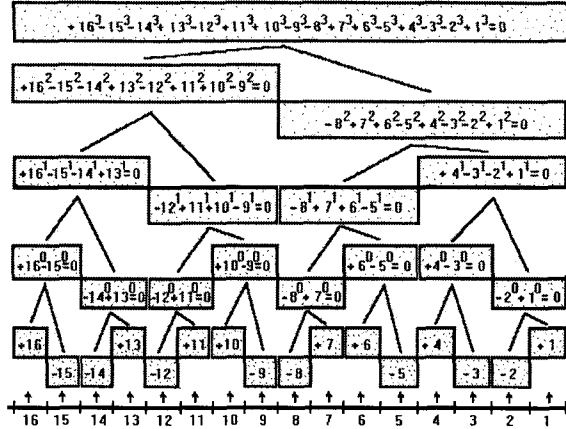
Figure 3. One of the hierarchical structures associated with the integer relations (7) and inequality (8). The structure may be interpreted as a result of a self-organization process of prime integer relations. It starts with integers $16, 13, 11, 10, 7, 6, 4, 1$ generated in the positive state and integers $15, 14, 12, 9, 8, 5, 3, 2$ generated in the negative state on the zero level. The self-organization process makes the prime integer relations from the first level to the fourth guided by a single principle. Arithmetic forbids the process from reaching level 5, as the left side of (8) cannot be zero.

4. Self-Organization Processes of Prime Integer Relations and Correlation Structures of Complex Systems

By using the dynamics $\Delta x_i, i = 1, ..., N$ of the elementary parts, the analysis of linear equations (5) and inequality (6) reveals hierarchical structures of prime integer relations in the description of a complex system. Through interpretation of such hierarchical structures we identify self-organization processes of prime integer relations [2], [3]. Starting with the integers as the elementary building blocks and controlled by arithmetic the self-organization processes make the prime integer relations of one level from the prime integer relations of the previous level.
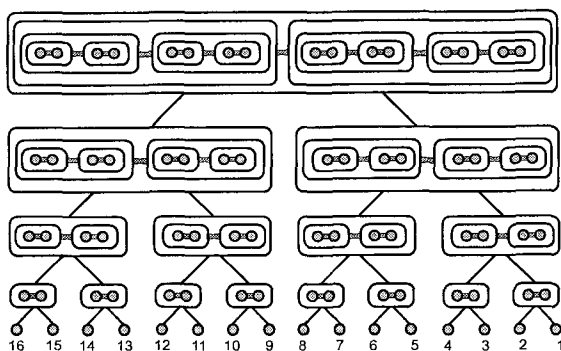
Figure 4. A correlation structure of the complex system is built in accordance with the hierarchical structure of prime integer relations (Figure 3). The elementary parts $P_{16}, ..., P_1$ are depicted as integers $16, ..., 1$ to show the correspondence between the composite parts of the complex system and the prime integer relations. The horizontal edge denotes that the parts of the complex system are correlated through a prime integer relation.

We illustrate the results by considering two states of a complex system

$$x = -1 + 1 + 1 - 1 + 1 - 1 - 1 + 1 + 1 - 1 - 1 + 1 - 1 + 1 + 1 - 1,$$

$$x' = +1 - 1 - 1 + 1 - 1 + 1 + 1 - 1 - 1 + 1 + 1 - 1 + 1 - 1 - 1 + 1$$

specified by the Prouhet-Thue-Morse (PTM) sequences of length $N = 16$. In this case the equations (5) become four integer relations $k = 0, 1, 2, 3$ as $C(x, x') = 4$ and $m = 0$

$$+16^0 - 15^0 - 14^0 + 13^0 - 12^0 + 11^0 + 10^0 - 9^0 - 8^0 + 7^0 + 6^0 - 5^0 + 4^0 - 3^0 - 2^0 + 1^0 = 0,$$

$$+16^1 - 15^1 - 14^1 + 13^1 - 12^1 + 11^1 + 10^1 - 9^1 - 8^1 + 7^1 + 6^1 - 5^1 + 4^1 - 3^1 - 2^1 + 1^1 = 0,$$

$$+16^2 - 15^2 - 14^2 + 13^2 - 12^2 + 11^2 + 10^2 - 9^2 - 8^2 + 7^2 + 6^2 - 5^2 + 4^2 - 3^2 - 2^2 + 1^2 = 0,$$

$$+16^3 - 15^3 - 14^3 + 13^3 - 12^3 + 11^3 + 10^3 - 9^3 - 8^3 + 7^3 + 6^3 - 5^3 + 4^3 - 3^3 - 2^3 + 1^3 = 0 \quad (7)$$

and the inequality (6) takes the form

$$+16^4-15^4-14^4+13^4-12^4+11^4+10^4-9^4-8^4+7^4+6^4-5^4+4^4-3^4-2^4+1^4 \neq 0, \quad (8)$$

where a common factor 2 originated from space-time variables $\Delta x_i, i = 1, ..., 16$ is not shown for clarity.
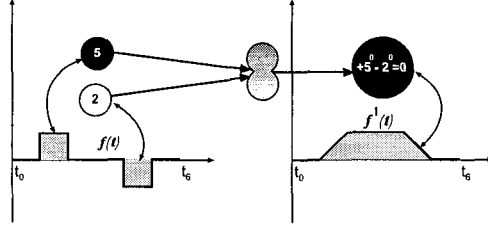
Figure 5. The integers and prime integer relation are pictured like "particles" to promote the integers and prime integer relations as objects of formation. Under the integration of the function, the geometric patterns corresponding to integer 5 and integer 2 form the geometric pattern corresponding to prime integer relation $+5^0 - 2^0 = 0$.

One of the hierarchical structures associated with the integer relations (7) and inequality (8) is shown in Figure 3. In the structure the relationships between the elements of neighboring levels may be interpreted as a consequence of a self-organization process of prime integer relations. The process starts with integers $16, ..., 1$ and proceeds level by level following the same organizing principle:

on each level the powers of the integers in the prime integer relations are increased by 1, so that through emerging interdependencies the prime integer relations could self-organize to form prime integer relations of the next level.

Remarkably, the formation of prime integer relations is more than their simple sum. Let us explain the notion of prime integer relation. A prime integer relation of the first level is the sum of integers from the zero level (Figure 3). Following the organizing principle a prime integer relation of one level is made from prime integer relations of the previous level. Notably, a prime integer relation is an indivisible element in the sense that it is not the sum of prime integer relations from the same level. For example, integer relation $+7^2 - 6^2 - 5^2 + 3^2 + 2^2 - 1^2 = 0$ is a prime integer relation, because none of its components can separately produce a prime integer relation. At the same time integer relation $+7^0 - 6^0 - 5^0 + 3^0 = 0$ is not prime, because it consists of two prime integer relations $+7^0 - 6^0 = 0$, $-5^0 + 3^0 = 0$.

The space-time dynamics Δx_i of elementary part $P_i, i = 1, ..., N$ has the following interpretation. The absolute value of Δx_i is the number of integers $(m+N-i+1)$ involved in a self-organization process associated with the equations (5) and inequality (6), while $sign(\Delta x_i)$ determines the state, i.e., positive or negative, of the integers, provided that $\Delta x_i \neq 0, i = 1, ..., N$.

The self-organization processes describe complex systems by hierarchical struc-

tures of prime integer relations. The prime integer relations are absolute true statements and thus provide information that cannot be further reduced. In particular, there exist no deeper principles to explain why the hierarchical structures exist the way they do and not otherwise. For example, it is not possible to explain further a structure, where prime integer relations $+7^0 - 6^0 = 0$, $-5^0 + 3^0 = 0$ and $+2^0 - 1^0 = 0$ of level 1 form a prime integer relation

$$+7^1 - 6^1 - 5^1 + 3^1 + 2^1 - 1^1 = 0$$

of level 2, which can self-organize itself into a prime integer relation

$$+7^2 - 6^2 - 5^2 + 3^2 + 2^2 - 1^2 = 0$$

of level 3. However, this prime integer relation cannot progress further to level 4, because

$$+7^3 - 6^3 - 5^3 + 3^3 + 2^3 - 1^3 \neq 0.$$

It is shown that the correlation structures underlying (3) and (4) are built in accordance with the hierarchical structures of prime integer relations associated with the equations (5) and inequality (6) [2], [5], [6]. For example, the prime integer

Figure 6.   A hierarchical structure of prime integer relations and an isomorphic hierarchical structure of geometric patterns. Due to the isomorphism the self-organization process of prime integer relations can be described geometrically and quantitatively. For example, we can see that geometric patterns on different levels are self-similar. A prime integer relation can be measured by the area of a corresponding geometric pattern.

relation $+4^0 - 3^0 = 0$ (Figure 3) corresponds to

$$+2 \cdot 4^0 - 2 \cdot 3^0 = \Delta x_4 4^0 + \Delta x_3 3^0 = 0,$$

which shows that space-time changes $\Delta x_4, \Delta x_3$ of the elementary parts P_4 and P_3 are correlated as $\Delta x_4 4^0 = -\Delta x_3 3^0$ and the parts form a composite part ($P_4 \leftrightarrow$

P_3). The correlation through the prime integer relation does not involve a physical signal. But, if the dynamics Δx_4 of the elementary part P_4 is specified, then the dynamics Δx_3 of the elementary part P_3 is instantaneously determined and vice versa. The correlation is nonlocal, because the prime integer relation does not have any reference to the distance between the parts. Similarly, the prime integer relation $-2^0 + 1^0 = 0$ specifies the correlation $\Delta x_2 2^0 = -\Delta x_1 1^0$ between the elementary parts P_2 and P_1 and a composite part $(P_2 \leftrightarrow P_1)$. In its turn the prime integer relation $+4^1 - 3^1 - 2^1 + 1^1 = 0$, made from prime integer relations $+4^0 - 3^0 = 0$ and $-2^0 + 1^0 = 0$, corresponds to

$$(\Delta x_4 4^1 + \Delta x_3 3^1) + (\Delta x_2 2^1 + \Delta x_1 1^1) = 0,$$

which shows that the composite parts $(P_4 \leftrightarrow P_3)$ and $(P_2 \leftrightarrow P_1)$ are correlated as

$$(\Delta x_4 4^1 + \Delta x_3 3^1) = -(\Delta x_2 2^1 + \Delta x_1 1^1)$$

and form a larger composite part $((P_4 \leftrightarrow P_3) \leftrightarrow (P_2 \leftrightarrow P_1))$.

In general, the self-organization process of prime integer relations can be associated with the formation of a correlation structure of the complex system (Figures 3 and 4). The formation process starts with the elementary parts $P_{16}, ..., P_1$ and combine them into composite parts in order to make larger composite parts and so on until the whole correlation structure is built (Figure 4).

By using the self-organization processes of prime integer relations a concept of complexity is introduced [2] and its applications may be found in [7].

5. Geometrization of the Self-Organization Processes and Fractals

In the above considerations prime integer relations appeared like objects of formation rather than objects of calculation. The prime integer relations might be better recognized as objects of formation, if they could be observed and measured as geometric patterns. Remarkably, by using the integer code series [3], it is shown that the prime integer relations can be geometrized as two-dimensional geometric patterns and the self-organization processes can be described in terms of their transformations [2]. These characterize the self-organization processes of prime integer relations geometrically and quantitatively. As a result we reveal fractal properties of the self-organization processes.

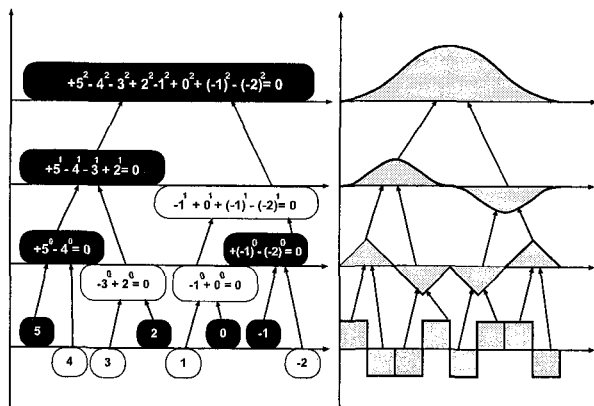Figure 5 illustrates the formation of a prime integer relation through the formation of a corresponding geometric pattern. A hierarchical structure of prime integer relations and an isomorphic hierarchical structure of geometric patterns are presented in Figure 6. The isomorphism allows us to observe fractal nature of the self-organization process of prime integer relations and investigate its scaling properties. A prime integer relation can be measured by the area of a corresponding geometric pattern.

In general, the geometric patterns of prime integer relations exhibit a variety of numbers with important properties. In particular, the PTM sequence is connected

with a self-organization process of prime integer relations progressing well through the hierarchical levels and demonstrating critical point [8], [9] features [2] (Figure 6). This resonates properly with the fact that the PTM sequence gives a symbolic description of chaos resulting from period-doubling [10] in physical systems [11]. Although the PTM geometric patterns at levels $n > 1$ are bounded by intricate curves, they have, nevertheless, a concise and universal description. For example, the area of a PTM geometric pattern at level $n \geq 1$ can be simply specified as $S = LH/2$, where L and H are the length and height of the geometric pattern (Figure 6). As a result the PTM geometric patterns have scale invariant properties dividing the hierarchical levels into similar groups of 3 successive levels.

Namely, the lengths and heights of the PTM geometric patterns at levels $n = 1, 2, 3$ and $n = 4, 5, 6$ are described in terms of ε and δ as $(2\varepsilon, \varepsilon\delta), (4\varepsilon, \varepsilon^2\delta), (8\varepsilon, 2\varepsilon^3\delta)$ and $(16\varepsilon, 8\varepsilon^4\delta), (32\varepsilon, 64\varepsilon^5\delta), (64\varepsilon, 1024\varepsilon^6\delta)$. By using the renomalization group transformation $\varepsilon' = 2^3\varepsilon, \delta' = \varepsilon^3\delta$, the lengths and heights at levels $n = 4, 5, 6$ can be expressed in terms of ε' and δ' in the same way $(2\varepsilon', \varepsilon'\delta'), (4\varepsilon', \varepsilon'^2\delta'), (8\varepsilon', 2\varepsilon'^3\delta')$ as the lengths and heights at levels $n = 1, 2, 3$ are given in terms of ε and δ. The situation repeats for levels $n = 7, 8, 9$ and so on.
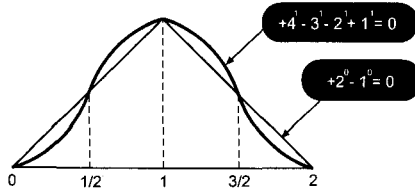
Figure 7. Two geometric patterns as resonant vibrations of "integer string". The first resonant vibration $v_1(t) = f^{[1]}(t)$ represents prime integer relation $+2^0 - 1^0 = 0$ and the second resonant vibration $v_2(t) = \frac{1}{\varepsilon}g^{[2]}(t'/2)$ represents prime integer relation $+4^1 - 3^1 - 2^1 + 1^1 = 0$, where $f = \rho_{m\varepsilon\delta}(x) \in W_{\varepsilon\delta}[t_0, t_2], x = +1 - 1, g = \rho_{m\varepsilon\delta}(x') \in W_{\varepsilon\delta}[t_0', t_4'], x' = +1 - 1 - 1 + 1$ and $m = 0, \varepsilon = 1, \delta = 1$.

In many cases it is useful to interpret the geometric patterns of prime integer relations as resonant vibrations of "integer string" (Figure 7).

6. Conclusions

We have shown that complex systems can be described in terms of self-organization processes of prime integer relations. The processes have the integers as the basic building blocks and under the control of arithmetic make the prime integer relations of one level from the prime integer relations of the previous level. Prime integer relations demonstrate important properties in the description of complex systems:

- May self-organize into another prime integer relations according to a single principle. A prime integer relation is more than the simple sum of the prime integer relations it is made of. A prime integer relation may participate in various self-

organization processes and as a result be a component of different prime integer relations at higher levels;

- Do not stand alone, but through the self-organization processes are all interconnected and inseparable in one hierarchical network. It seems that not a minor change can be made to any element of the network;

- Provide nonlocal and instantaneous correlations between the parts of a complex system;

- Can be described as two-dimensional geometric patterns and may be quantified in renormalizable numbers;

- Provide a complexity order. The self-organization processes start with different integers and as a result progress to different levels producing the complexity order. It seems that a self-organization process aims to progress as higher as possible in the direction of this order.

Finally, we have revealed the importance of prime integer relations as new elements in dealing with complex systems. The self-organization processes of prime integer relations describe complex systems by absolutely true information and cannot be explained in terms of deeper principles. This gives the possibility to develop an irreducible theory of complex systems.

References

1. L. Smolin, *Three Roads to Quantum Gravity* (Basic Books, 2001).
2. V. Korotkikh, *A Mathematical Structure for Emergent Computation* (Kluwer, Dordrecht, 1999).
3. V. Korotkikh, *Integers: Irreducible Guides in the Search for a Unified Theory*, Braz. J. Phys, Special Issue on Decoherence, Information, Complexity and Entropy, **35**(2B), 509 (2005).
4. V. Korotkikh, *Integer Code Series with Some Applications in Dynamical Systems and Complexity* (Computing Centre of RAS, Moscow, 1993).
5. G. Korotkikh and V. Korotkikh, in *Optimization and Industry: New Frontiers*, ed. by P. Pardalos and V. Korotkikh (Kluwer, Dordrecht, 2003).
6. V. Korotkikh, in *Fuzzy Partial Differential Equations and Relational Equations*, ed. by M. Nikravesh, L. Zadeh and V. Korotkikh (Springer, Berlin, 2004).
7. V. Korotkikh, G. Korotkikh and D. Bond, *On Optimality Condition of Complex Systems: Computational Evidence*, arXiv:cs/0504092, April, 2005.
8. L. Kadanoff, W. Gotze, D. Hamblen, R. Hecht, E.A.S. Lewis, V.V. Palciaukas, M. Rayl and J. Swift, Rev. Mod. Phys. **39**(2), 395 (1967).
9. K. G. Wilson and J. Kogut, Phys. Rep. C**12**, 75 (1974).
10. M. Feigenbaum, Los Alamos Sci. **1**, 4 (1980).
11. J. Allouche and M. Cosnard, in *Dynamical Systems and Cellular Automata* (Academic Press, 1985).

DYNAMICAL DECOMPOSITION OF MULTIFRACTAL TIME SERIES AS FRACTAL EVOLUTION AND LONG-TERM CYCLES: APPLICATIONS TO FOREIGN CURRENCY EXCHANGE MARKET

A. TURIEL

Institut de Ciencies del Mar- CMIMA (CSIC)
Passeig Maritim de la Barceloneta, 37-49.
08003 Barcelona. Spain
E-mail: turiel@icm.csic.es

C. PEREZ-VICENTE

Grup de Sistemes Complexos.
Departament de Fisica Fonamental.
Universitat de Barcelona. Diagonal, 647.
Barcelona 08028.Spain.
E-mail: conrad@ffn.ub.es

The application of the multifractal formalism to the study of some time series with scale invariant evolution has given rise to a rich framework of models and processing tools for the analysis of these signals. The formalism has been successfully exploited in different ways and with different goals: to obtain the effective variables governing the evolution of the series, to predict its future evolution, to estimate in which regime the series are, etc. In this paper, we discuss on the capabilities of a new, powerful processing tool, namely the computation of dynamical sources. With the aid of the source field, we will separate the fast, chaotic dynamics defined by the multifractal structure from a new, so-far unknown slow dynamics which concerns long cycles in the series. We discuss the results on the perspective of detection of sharp dynamic changes and forecasting.

1. Introduction

The study of the statistical properties of economic time series is the focus of intense research in the last years [1,2]. There is a continuous development of new methodological tools for the characterization and processing of econometric variables. This process is very important because it allows to create models to mimic the evolution of real systems and to identify the relevant mechanisms which govern their dynamics.

Many time series of variables with econometric interest (as for instance stock market price series, option rates or foreign currency exchange rates) are known to exhibit multifractal properties [3,4,5,6,7,8]. Apart from the typical statistical properties of scale and translational invariance the main feature of multifractal systems

is that they can be decomposed into a collection of sets (the fractal components) arranged in a hierarchical manner. Under changes in scale a multifractal behaves in a non-trivial way, due to the combination of scalings of each fractal component, and leading to the emergence of a collective property known as multiscaling. Almost since the origins of the multifractal formalism, two basic different approaches have been established to deal with multifractal systems. One option is to carry out an statistical analysis on the symmetries of the system, the other is to investigate the geometric aspects derived from the decomposition of the system in fractal sets.

Historically speaking the first approach is the most frequently used. It has allowed to develop a rich theory about multifractals, relating them with other properties, as energy injection cascades, intermittency and anomalous scaling (see [9] for a discussion). Nevertheless it has some difficulties. Statistical approaches are very demanding in data (multiscaling distributions are highly curtotic) and, in addition, there is no intuition about the properties of series at any specific point. Some of these problems can be solved by using a geometric approach. Methods of this type are more technical since they require an efficient decomposition of the series in the different fractal components but they are more powerful and have many advantages, as we will discuss throughout the paper. Here, we will apply the geometrical approach to the analysis of foreign currency market data series. In particular, emphasis is put on the most singular fractal component (from which the original series can be fully reconstructed) as well as on the concept of source, that allows to identify different dynamical regimes with different physical (economic) meaning.

2. Reconstructible multifractals

Let us start by giving a short theoretical introduction. To analyze a given temporal series denoted by $s(t)$ one defines a measure μ of any set A, as

$$\mu(A) \equiv \int_A d\tau \, |s'|(\tau) \tag{1}$$

where $s'(t)$ stands for the derivative of $s(t)$. The quantity $\mu(A)$ is a measure of the variations in the series $s(t)$ across the set A. The measure μ allows to locally characterize the scale invariance of the fractal components. We will say that a time series $s(t)$ is multifractal if for any Δt and any time instant t_0

$$\mu\left([t_0 - \Delta t/2, t_0 + \Delta t/2]\right) = \alpha_0 \Delta t^{d+h_0} + O(\Delta t^{d+h_0}) \tag{2}$$

where h_0 is a local exponent depending on the point t_0 considered and the shift d in the scaling has been introduced by normalization convenience; it corresponds to the dimension of the space of variables ($d = 1$ in our case)[10,11]. When a measure μ is multifractal, a scaling exponent h_0 can be assigned to every point t_0 which defines how changes of scale are locally performed. These local scaling exponents are referred to as *singularity exponents* because they characterize the functional

character -*i.e.* smooth, continuous, singular...- of $s(t)$ at each one of its points. Any technique assigning a singularity exponent to every point in the series is called *singularity analysis*. The obtention of the singularity exponents at each point of an experimental, discretized series require of an appropriate interpolation scheme and of some filtering capabilities, all that usually attained by the use of wavelet projections[12,13,11,10].

Once singularity exponents are known at every time t (which we will represent by the function $h(t)$), it is possible to classify points according to common values of the exponents. We define the fractal component F_h associated to the singularity exponent h as the set of points verifying $F_h \equiv \{h(t) = h\}$; thus, those points for which local changes in scale are performed in the same way belong to the same fractal component. Once F_h are obtained, their fractal dimensions $D(h)$ can be calculated. This function, known as the *singularity spectrum* of the multifractal, plays a central role in the connection of geometrical and statistical properties of multifractals: the famous Parisi-Frisch's formula[14] express the multiscaling exponents obtained in the statistical approach in terms of $D(h)$. Hence, series can be classified according to $D(h)$.

The singularity spectrum of a physical multifractal signal is always bounded from below, that is, there always exists a finite minimum singularity exponent h_∞ since physical signals cannot display divergences to infinity. The existence of a lower bound for the singularity exponents have deep consequences for natural multifractals, as the related fractal manifold is very informative about the signal. We will call the fractal manifold associated to the most singular (*i.e.*, smallest) exponent h_∞ the Most Singular Manifold (MSM) of the multifractal hierarchy, and we will denote it by F_∞ (*i.e.*, $F_\infty \equiv F_{h_\infty}$). The MSM corresponds to rare events in the distribution of singularities, located at the tail of that distribution. But in spite of being unlike events, the dynamics of the series is lead by the MSM[10]. In fact, based on the statistical interpretation of the MSM[16] it has been proposed that multifractal signals with lower bounded spectra can be reconstructed from the values of the gradient of the signal over the MSM. An explicit reconstruction formula was derived[16], that for the case of 1D series takes the following form[10]:

$$s(t) = s(0) + \int_0^t d\tau \, s'(\tau) \tag{3}$$

Equation (3) must be understood as an integral with respect to the Hausdorff measure associated to the MSM (we assume that the MSM is a regular fractal[15]) of the values of the gradient of s restricted to the MSM. Although when this expression is applied to real, discretized signal it gives a stepwise approximation to the data, in theory more complex behaviours could be obtained from it. One of the consequences of the existence of MSM and its relevance in terms of reconstruction is that multifractal series are structured objects, containing points which convey more information than others. In addition, the series can be described in a more

simplified way since the whole set of events can be coded in terms of the MSM, very useful for compression purposes [16].

3. Data. Performance of the reconstruction algorithm

For this work, we have made use of several series obtained from the foreign currency exchange market for some important currencies, sampled at a daily rate starting in January 1st, 1992 until December 31st, 1999. All the illustrations in this paper have been done for the Euro-US Dollar exchange rate. We consider the logarithm of the exchange rate, because in such a way the gradients approximate the relative variation in the exchange rate, what is the econometrically relevant quantity.

As a first step, we computed the singularity exponents associated to each time instant in the series by means of wavelet projections; see the details in[10]. When defining the fractal components on real data, we should accept an experimental quantization in the singularity determination Δh, which express the experimental uncertainty in the value. Previous works[11,10] show that quantizations of order $\Delta h = 0.1$ are acceptable for multifractal decomposition. For reconstruction task, a more exigent choice is necessary in order to attain reasonable reconstructions qualities. For the experiences in this paper, we fix a rather coarse determination of the MSM comprising the 30% of the singular points; such a amount guarantees good reconstruction qualities and efficient analysis of the series structure.

4. Reduced series

The MSM can be employed to generate a new type of signal, which is only concerned with the presence of the MSM: the reduced signal. The reduced signal $r(t)$ is generated when the reconstruction formula is applied to a simplified version of the derivatives: only the sign is kept. Hence, we take $r'(t)_{|F_\infty} = \text{sign}(s'(t))$ and we assume that r is a multifractal reconstructible signal, so we apply Eq. (3) to retrieve $r(t)$. Once the reduced series is constructed, we can observe that there is a striking similarity between the original and the reduced series. Compare Figs. 1 and 2, left: a visual inspection of the graphs reveal that the reduced and the original series have locally the same shape, just different on the size or amplitude of the observed structures. This correspondence reaches further: the equivalence of multifractal structures for both series can be theoretically granted, assuming that the values of $|s'|$ change smoothly over the MSM[17]. Nevertheless, there are points where sudden differences are observed, and for which multifractal structures do not coincide. To understand the nature of these transitions points and to characterize them in a systematic way we need to introduce the concept of *source*.

5. Sources

Sources are introduced in order to provide a consistent theoretical background in support of the hypothesized continuous function which seems to relate the series s
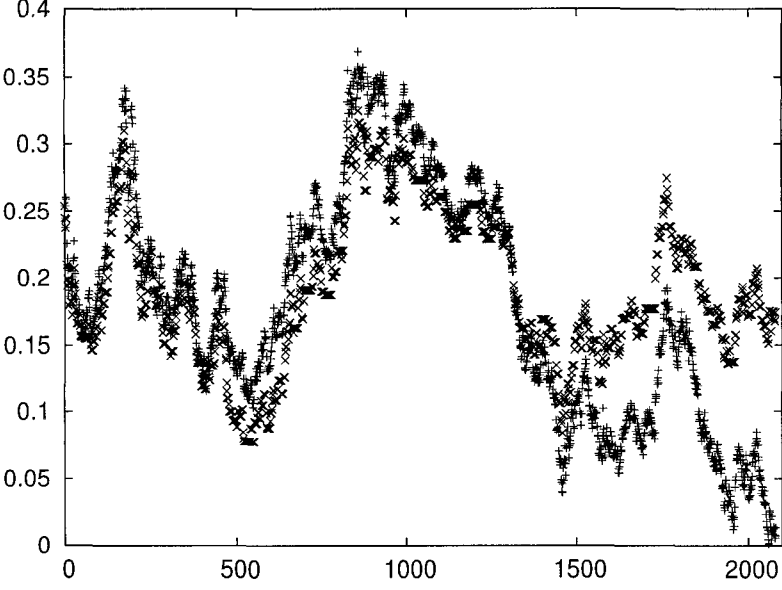
Figure 1. Original series of the logarithms of EUR-USD currency exchange rates (+) and its reconstruction from the MSM at 30% (x). The series is reconstructed in accumulated error, Eq. 3, and for that reason error increases with time.

and its reduced counterpart r. We look for a function $\rho(t)$ (that we will call the source function) which is continuous and relates the derivatives of the original series s and its reduced counterpart r, in the way:

$$s'(t) = \rho(t) r'(t) \tag{4}$$

Equation (4) has a crucial importance, as it gives an explicit representation of the original series as a combination of sources and reduced series. We can integrate both sides of the equation to obtain a new reconstruction formula, in the way:

$$s(t) = s(0) + \int_0^t d\tau\, \rho(\tau)\, r'(\tau) \tag{5}$$

This equation is in fact a generalization of the MSM-based reconstruction formula, Eq. (3), in which we have made explicit the separation between the continuous and the multifractal parts. The question is now how to obtain the source function $\rho(t)$. If we assume that both the signal $s(t)$ and the reduced series $r(t)$ define multifractal measures (denoted μ_s and μ_r) with the same singularity exponents, Eq. (5) can be interpreted in terms of Riesz's representation theorem[18], in which the measure μ_s is represented in terms of μ_r: $\mu_s(A) = \int_A d\mu_r(\tau)\, \rho(\tau)\ \forall A$. We

can evaluate the integral density $\rho(t)$ as the Radon-Nikodym derivative of μ_s with respect to μ_r. Notice that if μ_s and μ_r have the same singularity exponents then $\mu_s([t - \Delta t/2, t + \Delta t/2]) \sim \Delta t^{d+h(t)} \sim \mu_r([t - \Delta t/2, t + \Delta t/2])$ and so both measures are absolutely continuous respect to each other [18]. The density $\rho(t)$ can hence be obtained by taking sequences of intervals of diminishing sizes,

$$\rho(t) = \lim_{\Delta t \to 0} \frac{\mu_s([t - \Delta t/2, t + \Delta t/2])}{\mu_r([t - \Delta t/2, t + \Delta t/2])} \tag{6}$$

The density $\rho(t)$ is called the Radon-Nikodym derivative of μ_s with respect to μ_r (sometimes represented by the symbol $\frac{d\mu_s}{d\mu_r}$), and we can identify it with the source field: Equation (6) yields Eq. (4) when the derivatives of s and r are continuous. Unfortunately, a direct application of Eq. (6) over experimental data is not easy and demands special care[17]. In a previous study over stock market data [17] the best method to evaluate the source field leads to hypothetize that the sources form a stepwise function. In this way, one relaxes our initial assumption of continuity on ρ to piecewise continuity. The piecewise continuous model we propose here is the simplest one, namely piecewise linearity: for each *source domain A* we have

$$s(t) = a_A r(t) + b_A \quad \forall t \in A \tag{7}$$

what implies

$$\rho(t) = a_A \quad \forall t \in A \tag{8}$$

so that sources are stepwise (constant over each region; see[17] for a description on the iterative method to retrieve the source domains). There are many advantages respect to other strategies. It gives access to more regular source function ρ by definition (*i.e.*, they are piecewise constant). It gives more importance to regions than to points. The stepwise method is stable with respect to discretization. When the MSM is calculated at different densities (that is, with different quantizations), the associated stepwise sources have a very similar aspect, i.e., stepwise method is stable with respect to the quality in the determination of the MSM. Moreover piecewise constant sources grant the correspondence of multifractal structures of s and r: indeed, $\mu_s = a_A \mu_r$ over A, so the singularity exponents coincide. In Figure 2, right, we show the sources calculated for our example series. Finally, the quality of the reconstructions obtained by combining the reduced series and stepwise sources is very good, even using a small number of source domains $\{A_i\}_{i=1,...,N}$. The results are shown in Figure 3.

When the sources are described in terms of piecewise continuous functions the series introduces some transition points (*i.e.*, the limits of the source domains) in which sharp changes take place. For the rest of points, data series are just a rescaled version of their reduced counterparts. Henceforth, the frontiers of source
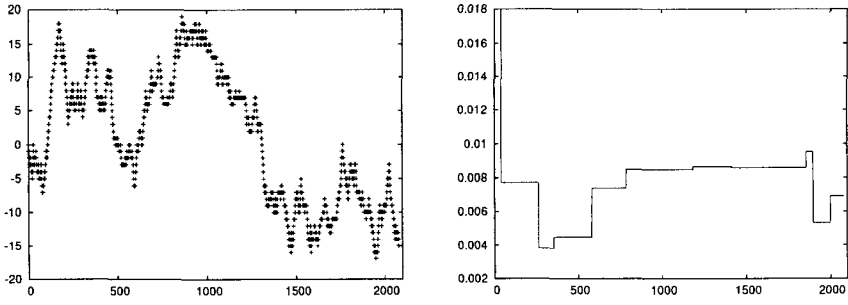
Figure 2. **Left:** Reduced series ; **Right:** Stepwise sources for EUR-USD daily series (with $N = 10$ temporal domains)
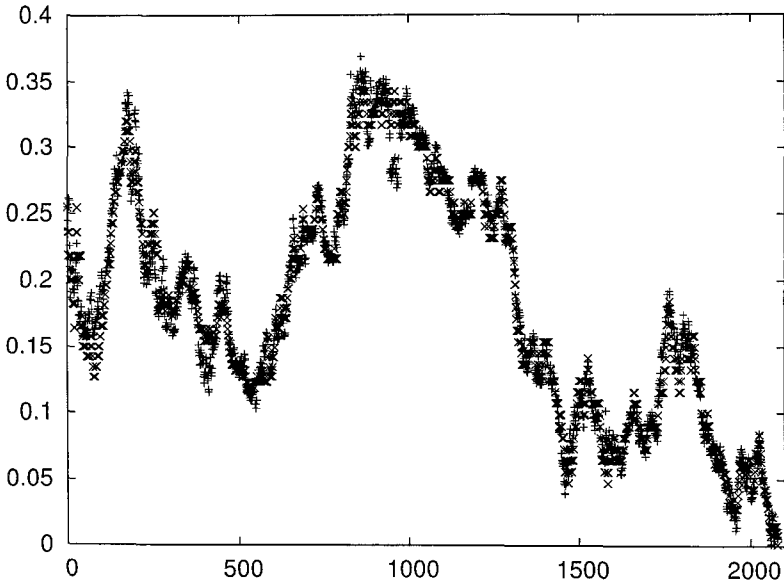
Figure 3. Original series of the logarithms of EUR-USD currency exchange rates (+) and its reconstruction applying sources (×); PSNR: 28.69 dB

domains have a dynamical interpretation: they are the transition points at which the economical system changes its behaviour. Those points are probably outside of the multifractal paradigm, as we can identify the multifractal structures of the original and the reduced series except for those points. Transition points in sources can probably be interpreted as too singular points in which statistics of temporal series deviates from multifractality, so they should be studied separately. Transition points are very relevant in order to understand changes in the dynamics of the

system; for instance, in other contexts as for instance meteorology the presence of transition source points have been related with the presence of convection currents, precipitations and other effects related with strong energy exchanges [19]. In our context, source plateaus are very close in value and transitions have small amplitudes, in contrast to what is observed in the study of stock market series [17]. A possible interpretation of source values and its transitions is the following. Let us consider the constant value a_A at each source domain A. By definition (see Eq. (4)) the absolute values of the derivatives of the series over the MSM equal that constant. The constant a_A represents then the maximum expected change per unit of time during the whole domain A. In this sense, a_A gives a measure of the maximum volatility[a] over that period: the volatility $\sigma(t_0, \Delta t)$ at a given time t_0 and over a window length Δt will be given by:

$$\sigma(t_0, \Delta t) \;=\; a_A \, \sigma_r(t_0, \Delta t) \tag{9}$$

where $\sigma_r(t_0, \Delta t)$ is the volatility for the reduced signal, that we will call reduced volatility. Taking into account that at any time step the reduced series can change at most by 1, $\sigma_r(t_0, \Delta t) < \Delta t$. According to our model, it follows that at each source domain $\sigma(t_0, \Delta t)/\Delta t < a_A$, which is consistent with the interpretation of a_A as maximum volatility per unit of time given above.

In spite of the simplicity of the reduced signal, in our model the volatility is a complicated, random variable because the reduced volatility itself is random. The reduced volatility depends on local variations in tendency (which depend on the particular basis point t_0 as well as on the size of the window extent Δt), on the Markov parameters characterizing the reduced series [10] (which in turn depend on the particular source domain A containing t_0) and in the multifractal structure (which does not evolve but may depend on the particular series we are studying; however this later dependency is integrated in the Markov parameters and does not require a separate study). An interesting prediction of our model is that all cycles and variations of volatility are also cycles and variations of reduced volatility, as both volatilities just differ in piecewise constant scale factors. In Figure 4 we observe that such prediction is correct over our example series. Notice that the very similar values in the source plateaus observed for the Euro-US Dollar exchange rate would imply that this market has a rather constant distribution of volatilities.

6. Conclusions and future work

The experimental evidence confirms that relevant econometric variables, such as volatility, can be characterized over long periods (the source domains) by the reduced series, which just contains geometrical information. The results confirm the

---

[a]Here we define the volatility at a given point as the standard deviation of values of series for a window of fixed size around that point.
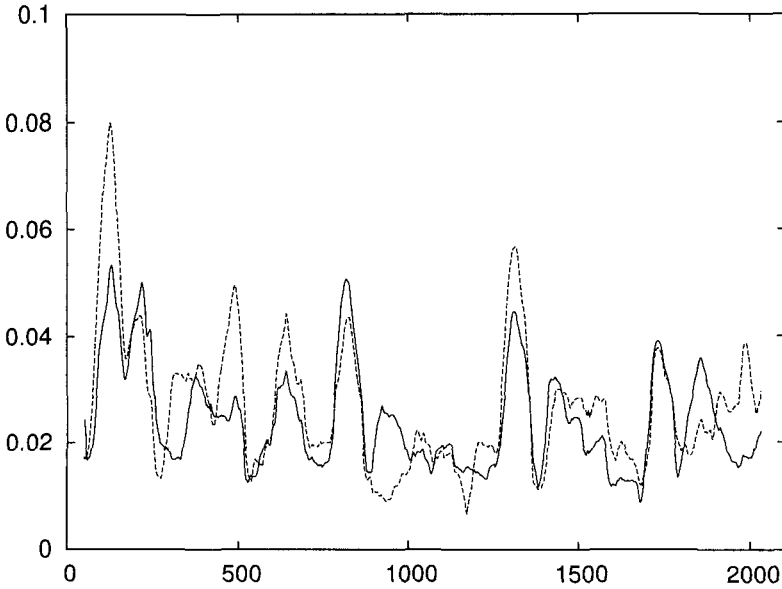
Figure 4. Volatility time series for the original data (continuous line) and the reduced counterpart (re-scaled by a factor 0.01 to fit the graph; dashed); $\Delta t = 100$ ticks. Notice that we can identify the same cycles for both volatility series, in agreement with theory.

importance of the MSM, as this set characterizes completely the reduced series. Any model for the volatility should primarily consist of a model for the reduced series (as the one presented in [10]). The sources have an important role, also. According to our model, volatilities must be recalculated once a transition point is crossed: following Eq. (9), the constant relating the reduced volatility with the actual volatility changes after each transition - as it is the source term. This fact makes the detection of transition points crucial.

The results presented here are obtained on the basis of a *descriptive* analysis, that is, the structure and the properties we have isolated are obtained by analyzing the series *a posteriori*: when processing a time instant in the series, we take into account both the past and the future times. But there is nothing in the multifractal formalism which prevents the use of unilateral intervals to analyze local scalings as in Eq. (2): namely:

$$\mu([to - \Delta t/2, t_0]) = \alpha_0 \Delta t^{d+h_0} + O(\Delta t^{d+h_0}) \tag{10}$$

Preliminary results on this type of *prospective* analysis (*i.e.*, oriented to the prediction) shows that the determination of the multifractal structure is still possible, although the experimental uncertainties increase, specially in the detection of points in the MSM. Timely detection of the MSM have relatively important uncertainties,

which decay fast as time passes and the point under study is further in the past of the last accessible point.

Concerning prospective detection of source domains, a new algorithm needs to be define, as the one presented[17] is by construction strictly descriptive and the determination of source domains is made in a global basis and not locally. This is an important limitation of the algorithm which is not intrinsic to sources; in fact, sources are local, as they correspond to the Radon-Nikodym derivative. New stable, unilateral methods for the obtention of sources on experimental, discretized data need to be implemented.

Acknowledgments

References

1. J. P. Bouchaud and M. Potters, *Theory of Financial Risk*, Cambridge University Press, Cambridge (2000).
2. R. N. Mantegna and H. E. Stanley, *An introduction to Econophysics*, Cambridge University Press (1999).
3. R. N. Mantegna and H. E. Stanley, *Nature*, **376**, 46 (1995).
4. S. Galluccio, G. Caldarelli, M. Marsili, and Y.-C. Zhang, *Physica A*, **245**, 423 (1997).
5. B. B. Mandelbrot, A. Fisher, and L. Calvet, *Cowles Foundation Discussion Paper No. 1164*, (1997).
6. A. Arneodo, J.-F. Muzy, and D. Sornette, *European Physical Journal B*, **2**, 277 (1998).
7. F. Schmitt, D. Schertzer, and S. Lovejoy, *International Journal of Theoretical and Applied Finance*, **3**, 361 (2000).
8. J. F. Muzy, J. Delour, and E. Bacry, *Euro. Phys. Journal B*, **17**, 537 (2000).
9. U. Frisch, *Turbulence*, Cambridge Univ. Press, Cambridge MA (1995).
10. A. Turiel and C. Pérez-Vicente, *Physica A*, **322**, 629 (2003).
11. A. Turiel and N. Parga, *Neural Computation*, **12**, 763 (2000).
12. I. Daubechies, *Ten lectures on wavelets*, CBMS-NSF Series in Ap. Math. Capital City Press, Montpelier, Vermont (1992).
13. A. Arneodo, F. Argoul, E. Bacry, J. Elezgaray, and J. F. Muzy, *Ondelettes, multifractales et turbulence*, Diderot Editeur, Paris (1995).
14. G. Parisi and U. Frisch, in *Turbulence and Predictability in Geophysical Fluid Dynamics. Proc. Intl. School of Physics E. Fermi*, M. Ghil, R. Benzi, and G. Parisi, Eds., North Holland, Amsterdam, 84 (1985).
15. K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley and sons, Chichester (1990)
16. A. Turiel and A. del Pozo, *IEEE Trans. on Im. Proc.*, **11**, 345 (2002).
17. A. Turiel and C. Pérez-Vicente, *Physica A*, **355**, 475 (2005).
18. W. Rudin, *Real and Complex Analysis*, Mc Graw Hill, New York (1987).
19. A. Turiel, J. Grazzini and H. Yahia to appear in *IEEE Letters on Geoscience and Remote Sensing*, (2005).

FRACTAL SETS FROM NONINVERTIBLE MAPS

CHRISTIAN MIRA
19 rue d'Occitanie, Fonsegrives, 31130 QUINT,
and Istituto di Scienze Economiche, University of Urbino (Italy).
E-mail: c.mira@free.fr

This text identifies some fractal sets generated by noninvertible maps. Two topics are presented: the fractal "*box-within-a-box*" bifurcations structure (described in 1975), produced by a dim 1 unimodal map, and the fractalization of basin boundaries generated by plane noninvertible maps. Some bifurcations of fractalization are described, depending on the basin boundary situation with respect to the critical curve (locus of points having two coincident preimages), and to the nature of strange repellers.

Keywords: noninvertible map, fractal set, stability, basin, bifurcation.

1 Introduction

The dynamic systems considered in this paper are described by a " discrete" equation, the solution of which is a sequence of points (orbit, or discrete trajectory) determined by an initial point (initial condition). The model of many processes are of such a type, called *recurrence relationship* (or simply *recurrence*). The corresponding explicit form is written:

$$X_{n+1} = F(X_n, \Lambda), \quad n = 0, 1, 2, \ldots \qquad X(n = 0) = X_0 \qquad (1)$$

Here X_n is a real phase (or state) vector, dim $X = p$, Λ a real parameter vector, the sequence of integers n is related to a discrete time, X_0 is the initial condition, F a single-valued smooth function of X_n. Depending on the scientific field such an equation is also called *iteration*, or *map*, or *point-mapping*. In this text (1) will be called map, denoted T. Its equation is symbolically represented by $X_{n+1} = TX_n$, or without lower indices $X' = TX$.

The point X' (or X_{n+1}) is called *rank-one image* of X (or X_n). The map $X_{n+r} = F_r(X_n, \Lambda)$ deduced from (1) after r iterations is denoted T^r, $X_{n+r} = T^r X_n$, X_{n+r} being the *rank-r image* of X_n. Considering the inverse map, the point $X = T^{-1}X'$ belongs to the set of *rank-one preimages* of X'. When T is noninvertible this set may be either made up of several points, or of only one point, or void. Then the phase space (defined by X) is made up of open regions Z_i, each point of a region Z_i having i distinct rank-one real preimages, $i = 0$ corresponding to a region without any real preimage. A point X_n belongs to the set of *rank-r* preimages of X_{n+r}. The simplest illustration of noninvertible map is given by the quadratic one-dimensional example ($p = 1$): $x' = x^2 - \lambda$ (Myrberg's map), λ being a parameter. The inverse map T^{-1} is given by $x = \pm\sqrt{x' + \lambda}$. So the x-axis is made up of the intervals Z_2 ($x' > -\lambda$), Z_0 ($x' < -\lambda$). The boundary point $x' = -\lambda$, which has two merging rank-one preimages at $x = 0$, is called *critical point* (in the Julia-Fatou sense).

The solution of (1), for the initial condition $X(n = 0) = X_0$, is a sequence of points:

$$X_n = X(n, X_0, \Lambda), \qquad n = 1, 2, 3, \dots \dots \tag{2}$$

X_n is called *iterated sequence*, or *discrete phase trajectory*, or *orbit*. In general $X(n, X_0, \Lambda)$ is a non classical transcendental function, which only the qualitative methods of nonlinear dynamics can characterize. In this framework a meaningful characterization consists in the identification of the $X(n, X_0, \Lambda)$ *singularities*, and the behavior changes of the latter (*bifurcations*) when Λ varies. The simplest singularities are the zero-dimensional ones: *period k cycles*, denoted also *k-cycles*. A *k-cycle* is a set of k consecutive points X_i^*, $i = 1, 2, ..., k$, permuting by successive applications of T, such that: $X_i^* = T^k X_i^*$, with $X_i^* \neq T^r X_i^*$ for $1 \leq r < k$. When $k = 1$ the point X^* is called a *fixed point* (period one cycle). A cycle is characterized by p *multipliers* S_j, $j = 1, ..., p$, which are the eigen-values of the associated Jacobian matrix. It is locally attracting (asymptotically stable) iff all $|S_j| < 1$. Only one $|S_j| < 1$ is sufficient to make repelling (or unstable) a cycle. When $|S_l| < 1$, $l = 1, ..., m$, $|S_r| < 1$, $r = m + 1, ..., p$, $m = 1, ..., p - 1$, the cycle is a *saddle*. When all $|S_j| > 1$ it said *expanding*. A period-k cycle is identified by the symbolism $(k; j)$, j being an index which characterizes the permutation of the k-cycle points by k successive applications of T. This index permit to differentiate cycles having the same period k and issued from different bifurcations (two cycles coming from the same bifurcation have the same points permutation). When $p > 1$ other classical singularities are stable and unstable manifolds of saddle cycles, basin boundaries, various invariant sets [1][2].

The text matter is devoted to fractal sets generated by one and two-dimensional noninvertible maps ($p = 1, 2$). Noninvertible maps generate the classical singularities of the invertible case (the inverse exists and is unique), i.e. fixed points, period-k cycles, $k = 2, 3, ...$, when $p \geq 1$, stable and unstable manifolds of a saddle point when $p > 1$. Nevertheless such maps T give rise to a new type of singularity the *critical set* CS (point if $p = 1$, curve if $p = 2$) the branches of which (except some particular cases) are the boundaries of the open Z_i regions. Noninvertible maps T can be identified according to the organization of the regions Z_i. The simplest maps are of $(Z_0 - Z_2)$ type (the space is divided into two unbounded regions Z_0 and Z_2 separated by CS), and $(Z_1 - Z_3 - Z_1)$ type (the space is divided into three unbounded regions: one Z_3, bordered by two regions Z_1). More complex types are described in [1]-[3]. The set CS is defined as the locus of points such that two determinations (rank-one preimages) of the inverse map T^{-1} are merging on a set CS_{-1}. When the map is smooth, CS_{-1} is obtained by equaling to zero the Jacobian determinant of T. The critical set plays a fundamental role in the solutions behavior and their bifurcations.

This paper essentially deals with two topics. The first one concerns the fractal "*box-within-a-box*", or "*embedded boxes*", bifurcations structure generated by a one-dimensional unimodal map (i.e. a $Z_0 - Z_2$ map), with some short informations on its extension to multi-modal maps (having intervals Z_i with $i > 1$) This structure was described for the first time from properties of self-similarity in 1975 [4] [5] before the introduction of the word "fractal". The second topic is related to the

basin boundary structure of an attracting set, when this boundary is fractal, in the case of plane noninvertible maps T ($p = 2$), i.e. CS is now a critical curve LC. For the maps class considered here, it is about to show some typical situations, and to explain the conditions of their birth with parameter variations giving rise to nonclassical bifurcations. As for nonfractal basin boundaries, these bifurcations are characterized by a common feature: *contact of the basin boundary with the critical set*. Even if the structure of LC is simple, the basin boundary is generally very complex with a large variety of geometrical situations, which leads to a large variety of contacts and fractalization features. The basins generated by plane noninvertible maps may be either *simply connected*, or *multiply connected*, or *non connected*, depending on the place of the basin boundary with respect to the critical curve LC [1] [2]. This dependence occurs in many different ways. It is reminded that the basin is multiply connected when it is pierced by holes, here called *lakes*. When it is nonconnected it is always made up of areas (here called *islands*) without any connection, the area containing the attractor being the *immediate basin*.

The fractal properties of a basin boundary are defined from a basic fractal core called *strange repeller SR*, which belongs to this boundary. It is constituted by a *nucleus* made up of infinitely many infinite sequences of unstable cycles with increasing period, their limit sets when the period tends toward infinity. The remaining part of SR consists of the increasing rank preimages (arborescent sequence) of all these points. When lakes, or islands, have a fractal structure, they are infinitely many and SR is their limit set. The knowledge of fractal properties of a basin boundary necessitates the identification of its related strange repeller, and the follow up of its birth conditions. The repeller SR may be located on a dim 1 manifold of the plane map T, considering the restriction (a noninvertible dim 1 map) T_r of T to this manifold. In this sense the bifurcation structure generated by T_r plays an important role in the SR identification when $p = 2$. The strange repeller also may come from the destruction of a chaotic attractor not located on a dim 1 manifold, it may be also generated from a perturbation of a Julia-Fatou set [2] [6], and for some maps families all these different types of strange repellers may coexist. It is worth noting that SR gives only the basic fractal properties of a basin boundary, a same type of SR leading to various types of basin boundaries. Then further information on the basin characteristics are obtained from the geometrical situation of the critical set LC with respect to a basin boundary.

Section 2 is a succinct presentation of the fractal "box-within-a-box" bifurcations structure generated by dim 1 unimodal maps, with the related symbolism. Section 3 is devoted to singularities and bifurcations specific to plane noninvertible maps. Two examples of basin boundary including a strange repeller SR of a dim 1 map are given in sec. 4. Section 5 presents an example for which SR comes from the destruction of a chaotic attractor. The case of a strange repeller resulting from the perturbation of the Julia-Fatou set of the complex map $z' = z^2 - \lambda$ is the sec. 5 object.

2 Box-within-a-box bifurcations structure of $\dim 1$ noninvertible maps

2.1 Rundown of basic properties

The "*box-within-a-box*" (translation of "*boîtes emboîtées*" in French) bifurcations structure is generated by noninvertible one-dimensional maps. Its first description was given in [4] [5] before the word "fractal" was introduced, the situation being described by the sentence "*the whole is similar to the parts even if they are infinitesimal*". Quoting these publications Guckenheimer called it "*embedded boxes*" [7]. After this structure has been also presented more completely in the books [2] [6] [8] [9]. It is characterized by the fact that the fractal organization of bifurcations is known by identification of non classical bifurcations resulting from the merging of two singularities of different nature (critical points and cycles points). The basic fractal bifurcation structure is generated by unimodal maps (i.e. defined by a function having only one extremum, a $Z_0 - Z_2$ map). The simplest case corresponds to the quadratic map,

$$T: \qquad x' = x^2 - \lambda \qquad\qquad (3)$$

Here x is a real variable, λ a real parameter. Such a map, called *Myrberg's map* [10] is defined by a parabola with an ordinate minimum C_{-1} ($x = 0$), its rank-one image $C = T(C_{-1})$ being the *rank-one critical point* of T. The inverse map T^{-1} is defined by $x = \pm\sqrt{x' + \lambda}$. The map T is characterized by the following properties.

- **(a)** All the bifurcations values of (3) occur into the parameter interval $\Omega_1 = [\lambda_{(1)_0}, \lambda_1^*]$, called *overall box*, $\lambda_{(1)_0} = -1/4$, $\lambda_1^* = 2$. For each λ-value, $-1/4 < \lambda < 2$, the map generates a unique attractor (a stable fixed point, or a stable period k cycle, or a chaotic attractor). The value $\lambda_{(1)_0} = -1/4$ corresponds to a *fold bifurcation* giving rise to two fixed points q_i, $i = 1, 2$, with multiplier $S = 2x(q_i)$: q_1 always unstable ($S > 1$), and q_2 ($S < 1$, attracting when $-1 \leq S < 1$). No real fixed point exists for $\lambda < \lambda_{(1)_0} = -1/4$. The value $\lambda = \lambda_1^* = 2$ is a basic non-classical bifurcation resulting from the merging of the unstable critical point q_1 with the rank-two critical point $C_1 = T(C) = T^2(C_{-1})$. It is characterized by the relations $x(C_1) = x(q_1) = 2$, $x(C) = x(q_1^{-1}) = -2$, $T^{-1}(q_1) = q_1 \cup q_1^{-1}$. When $0 < \lambda < \lambda_1^*$ the invariant segment $[q_1^{-1}, q_1]$ is the closure of the basin of the absorbing segment $\overline{CC_1}$, which contains the unique attractor. When $\lambda = \lambda_1^*$ the segment $\overline{CC_1}$ is chaotic and merges with $[q_1^{-1}, q_1]$, all the possible cycles have been created, they belong to $\overline{CC_1}$. Then $\overline{CC_1}$ is *invariant but not absorbing*. An initial condition x_0 on this segment gives a bounded orbit, belonging to the interval $[q_1^{-1}, q_1]$, orbit very sensitive to very small changes of x_0. The repelling cycles and their increasing rank preimages constitute a real set (E), occupying the whole interval $[-2 \leq x \leq 2]$, the derived set (set of limit points) (E') of which is perfect. The preimages of (E) are everywhere dense on (E') (see more details in [6] [8] [9]).

For $\lambda > 3/4$ the fixed point q_2 is always repelling with $S(q_2) < -1$ and a period-$k = 2$ cycle, made up of two points q_{2i}, $i = 1, 2$, having the multiplier $S(q_{2i}) = 4 - 4\lambda$, emerges from q_2. The value $\lambda = \lambda_{b1} = 3/4$ corresponds to a *flip bifurcation*. New increasing values of λ generates a sequence of flip bifurcations for $\lambda = \lambda_{bm}$ related to period 2^i cycles, $i = 1, 2, ..., m$, with an accumulation value

$\lim_{m\to\infty} \lambda_{bm} = \lambda_{1s} \simeq 1.401155189$ [10]. For $\lambda < \lambda_{1s}$ the number of repelling cycles is finite. They have the period 2^m, and have been created after crossing through each λ_{bm} value. For $\lambda > \lambda_{1s}$ infinitely many repelling period 2^i cycles $(i = 0, 1, 2, ...)$ exist, and they constitute a fractal set. Myrberg calls *spectrum* the parameter interval $\omega_1 \equiv [\lambda_{(1)_0}; \lambda_{1s}]$, here related to the sequence (cascade) of bifurcations by period doubling from the fixed point q_2 $(i = 0)$.

- **(b)** The number N_k of all possible cycles having the same period k, and the number $N_\lambda(k)$ of bifurcation values giving rise to these cycles, increase very rapidly with k (for more details cf. [9]). Cycles with the same period k differ from one another by the cyclic transfer (permutation) of one of their points by k successive iterations by T. Then a period k-cycle is identified by the symbolism $(k; j)$, j being an index characterizing this cyclic transfer. Considering one of such cycles, it can be generated from two basic bifurcations: either a fold one, or a flip one. A fold bifurcation generates two *basic cycles* at $\lambda = \lambda^j_{(k)_0}$: $(k; j)_{S>1}$ and $(k; j)_{S<1}$, $k \neq 2$. With increasing values of λ, from the cycle $(k; j)_{S<1}$ a flip bifurcations cascade occurs for $(k2^i; j_i)_{S<1}$ cycles with accumulation at a value λ^j_{ks} when $i \to \infty$, $2 > \lambda^j_{ks} > \lambda_{1s}$. Myrberg also calls "*spectrum*" the parameter interval $\omega^j_k = [\lambda^j_{(k)_0}; \lambda^j_{ks}]$, $k = 1, 3, 4, ...$, made up of parameter intervals leading to attracting cycles of period $k2^i$, $i = 0, 1, 2, ...$. The cycle symbolisms $(k; j)$ and $(k2^i; j_i)$ correspond to what is called a *nonembedded representation* in [2] [9].

When $\lambda > \lambda^*_1 = 2$, $[q_1^{-1}, q_1] \subset \overline{CC_1}$, the only attractor is the point at infinity, and no bifurcation takes place. The set (E) (without the point at infinity) constitutes the non-wandering set $(E) \in [q_{-1}^1; q_1]$. All the possible cycles have been created by (3). They are real and repelling, (E) is *completely disjointed*, and is a *Cantor set*. The situation equivalent to the λ^*_1 one (but now with an attractor inside $\overline{CC_1}$) is met for a $(k; j)$ cycle with multiplier $S > 1$ (thus generated by a fold bifurcation), for a parameter value denoted λ^{*j}_k. In this case λ^{*j}_k is the least λ-value such that the critical points $C_k = T^k(C)$, $C_{k+1}, ..., C_{2k-1}$ merge into the k points of the $(k; j)$ cycle with $S > 1$. By consideration of T^k the value λ^{*j}_k reproduces qualitatively the situation of T when $\lambda = \lambda^*_1$. Differently from $\lambda = \lambda^*_1$, for λ^{*j}_k the map T generates k attracting segments constituting a *period-k chaotic segment* which is now *absorbing*.

2.2 Description of the structure

The fractal "box-within-a-box" structure, here defined for the map (3), concerns all unimodal maps ($Z_0 - Z_2$ ones) with correctly chosen parameter variation. This structure is of "Russian dolls" type. It considers the parameter interval $\Omega_1 = [\lambda_{(1)_0}, \lambda^*_1]$ (*overall box*) inside which all the possible bifurcations occur. Out of Ω_1 no bifurcation happens. Ω_1 is generated from the two basic period $k = 1$ cycles, i.e. the fixed points q_1 and q_2. Taking into account the Myrberg' spectrum ω_1 related to the fixed point q_2 ($S < 1$), the box Ω_1 is defined by:

$$\Omega_1 = [\lambda_{(1)_0}, \lambda^*_1] = \omega_1 \cup \Delta_1 \qquad \Delta_1 =]\lambda_{1s}, \lambda^*_1]$$

Note that the symbolism "2^i" excludes even period cycles born from a fold bifurcation, or a flip bifurcation related to a *basic cycle* appearing out of ω_1 ($2^2 \neq 4$,

$2^3 \neq 8$, ..., with such a symbolism). So cycles different from $(2^i; j_i)$ can appear only for $\lambda \in \Delta_1$. For $\lambda < \lambda_{(1)_0} = -1/4$ the map has neither fixed point (except the point at infinity) nor cycles, every orbit is divergent. For $\lambda > \lambda_1^* = 2$ all the possible period k cycles have been created. They are repelling, and the map has the properties indicated in sec. 2.1.

Two period-k ($k = 3, 4, ...$) basic cycles $(k; j)$, one with $S > 1$, the other with $S < 1$, generate a parameter interval, provisionally denoted $\widehat{\Omega}_k$, having the same behavior as Ω_1, $\widehat{\Omega}_k \subset \Delta_1$. If $\widehat{\Omega}_k$ is not contained into another interval $\widehat{\Omega}_{k'}$, k being a multiple of k', $\widehat{\Omega}_k$ is called *rank-one box*, or *box* $(k; j)$ (*embedded representation*). It is denoted Ω_k^j with:

$$\Omega_k^j = [\lambda_{(k)_0}^j, \lambda_k^{*j}], \quad \Delta_k^j =]\lambda_{ks}^j, \lambda_k^{*j}], \quad \Omega_k^j = \omega_k^j \cup \Delta_k^j \subset \Delta_1$$

Here ω_k^j is the *spectrum* $(k; j)$. Then for T^k the box Ω_k^j reproduces all the bifurcations contained in the box Ω_1, in the same order (self similarity property), for a set of cycles having periods multiple of k. Denote $\Omega_{k_1}^{j_1}$ one of such boxes. Inside $\Omega_{k_1}^{j_1}$ it is possible to define *rank-two* boxes $\Omega_{k_1.k_2}^{j_1,j_2} = [\lambda_{(k_1.k_2)_0}^{j_1,j_2}, \lambda_{k_1.k_2}^{*j_1,j_2}] \subset \Delta_{k_1}^{j_1}$, related to two $(k_1.k_2; j_1, j_2)$ basic cycles, which for $(T^{k_1})^{k_2}$ undergoes in the same order all the bifurcations of the box $\Omega_{k_1}^{j_1}$, and so those of Ω_1:

$$\Omega_{k_1.k_2}^{j_1,j_2} = [\lambda_{(k_1.k_2)_0}^{j_1,j_2}, \lambda_{k_1.k_2}^{*j_1,j_2}] = \omega_{k_1.k_2}^{j_1,j_2} \cup \Delta_{k_1.k_2}^{j_1,j_2} \subset \Delta_{k_1}^{j_1}, \quad \Delta_{k_1.k_2}^{j_1,j_2} =]\lambda_{(k_1.k_2)s}^{j_1,j_2}, \lambda_{k_1.k_2}^{*j_1,j_2}]$$

More generally it is possible to define *rank-a boxes*, embedded into a rank-$(a-1)$ box

$$\Omega_{k_1....k_a}^{j_1...,j_a} = [\lambda_{(k_1....k_a)_0}^{j_1...,j_a}, \lambda_{k_1....k_a}^{*j_1...,j_a}] = \omega_{k_1....k_a}^{j_1...,j_a} \cup \Delta_{k_1....k_a}^{j_1...,j_a} \subset \Delta_{k_1....k_{a-1}}^{j_1...,j_{a-1}}$$

So $\Omega_{k_1....k_a}^{j_1...,j_a} \subset \Omega_{k_1....k_{a-1}}^{j_1...,j_{a-1}}$, $a = 1, 2, ...$. The boundary parameter λ^* of each of these boxes ($a = 1, 2, ...$) is defined from the the merging of well defined critical points with a repelling basic cycle having the multiplier $S > 1$. They are called *boxes of first kind*. The representation of these boxes is given by Fig. 1a, with the enlargement of Fig. 1b.

Other type of parameter λ^* can be also defined from a repelling cycle with $S < -1$, born from a flip bifurcation. Then when $\lambda = \lambda_{k2^i}^{*j}$, $i = 1, 2, ..., k = 1, 3, 4, ...$, $k2^{i-1}$ critical points of well defined rank merge into a period $k2^{i-1}$ cycle with $S < -1$ (born from a flip bifurcation). We call *box of second kind*, a box denoted $\Omega_{k2^i}^j$ which has $\lambda_{k2^i}^{*j}$ as one of the two boundary points (cf. Fig. 1c for $k = 1$). The second boundary point is the flip bifurcation $\lambda = \lambda_{bk2^{i-1}}^{ji}$, parameter generating the attracting cycle $(k; 2^i)$. When $i \to \infty$, the two boundaries tend toward λ_{ks}^j, with $\lambda_{k2^i}^{*j} > \lambda_{ks}^j$. The first and largest box of second kind is $\Omega_{2^1} \equiv [\lambda_{b1}, \lambda_{2^1}^*]$, $\lambda_{2^1}^*$ ($k = 1$) corresponding to $C_2 \equiv q_2$. Figure 1 represents the whole "box-within-a-box" (or embedded boxes) bifurcations structure. It is fractal because the boxes are self similar, i.e. the organization of the set Ω_1 is similar to that of its parts (the above defined boxes), even if these parts are infinitesimal.

Ω_1

Δ_1 ω_1

Myrberg's cascade $k=1$

Ω_3

Δ_3 ω_3

No period odd cycle

Ω_4^1 Ω_2^1

Ω_5^3 Ω_5^3 $\Omega_{3.3}^{1,1}$ Ω_5^1

λ

λ_1^* λ_3^* $\lambda_{(3)_0}$ λ_2^{*1} λ_{b1} $\lambda_{(1)_0}$

2 $\Omega_{3.4}^{1,1}$ λ_{3s} (a) $\lambda_{1s} \approx 1.401$ $-1/4$

(a)

$\Omega_{k_1}^{j_1} \subset \Delta_1$

$\Delta_{k_1}^{j_1}$ $\omega_{k_1}^{j_1}$

Myrberg's cascade $k=k_1$

$\Omega_{k_1.3}^{j_1,1}$ $\omega_{k_1.3}^{j_1,1}$

$\Omega_{k_1.4}^{j_1,1}$ $\Delta_{k_1.3}^{j_1,1}$ $\Omega_{k_1.5}^{j_1,1}$ $\Omega_{k_1.2^1}^{j_1}$

$\Omega_{k_1.5}^{j_1,3}$ $\Omega_{k_1.5}^{j_1,2}$ $\Omega_{k_1.3.3}^{j_1,1,1}$

$\lambda_{k_1}^{*j_1}$ $\lambda_{k_1.3}^{*j_1,1}$ $\lambda_{(k_1.3)_0}^{j_1}$ $\lambda_{k_1.2^1}^{*j_1}$ $\lambda_{k_1s}^{j_1}$ $\lambda_{k_1 b_1}^{j_1}$ $\lambda_{(k_1)_0}^{j_1}$

$\lambda_{k_1.3s}^{j_1,1}$

Limit of boxes Limit of boxes

$\Omega_{k''}^{j'},\ k'' \to \infty$ $\Omega_{k'}^{j},\ k' \to \infty$

(b)

Ω_{2^1}

Ω_{2^2}

$\Omega_{2^1.m}^{j_1}$ Ω_{2^i}

$\lambda_{2^1}^*$ $\lambda_{2^2}^*$ $\lambda_{2^i}^*$ λ_{1s} λ_{bi} λ_{b2} λ_{b1}

(c)
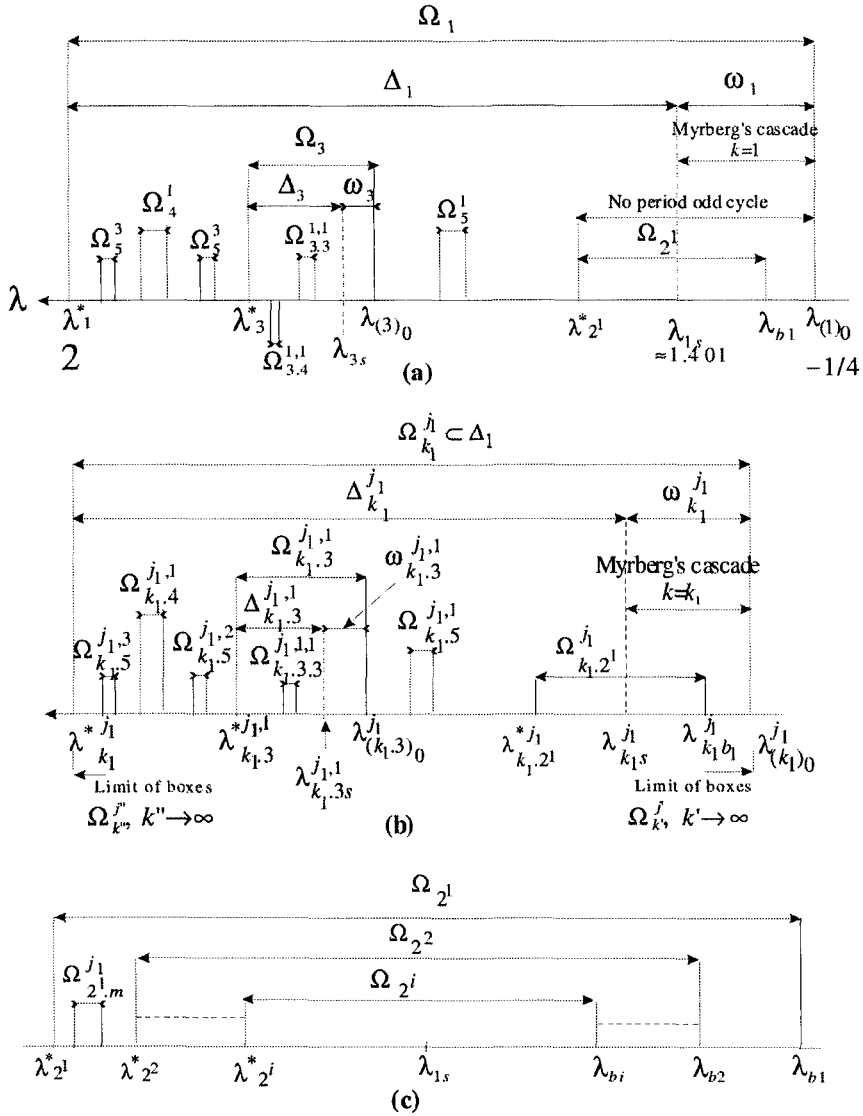
Figure 1. *Box-within-a-box* (or *embedded boxes*) bifurcations structure.

2.3 Properties

The box boundaries satisfy the following properties: out of Ω_k^j, $\lambda < \lambda_{(k)_0}^j$, the value $\lambda_{(k)_0}^j$ is a *limit point of boxes* $\Omega_{k'}^{j'}$ for $k' \to \infty$, with $\lambda > \lambda_k^{*j}$ the value λ_k^{*j} is a *limit point of boxes* $\Omega_{k''}^{j''}$ when $k'' \to \infty$. Considering the map (3) and increasing values of the parameter λ, the multiplier S of a cycle $(k;j)_{S>1}$ increases and the multiplier S of a cycle $(k;j)_{S<-1}$ decreases. So these cycles become more and more repelling,

and they cannot disappear by bifurcation. The following properties results:

- (a) *Let $[k, j]$ (nonembedded representation), $k = 1, 3, 4, ...$, be the given basic cycle of the box Ω_k^j with $S < 1$. For $\lambda \geq \lambda_{ks}^j$, Ω_k^j has generated a Cantor set $Cs[k, j]$ made up of all the repelling $(k2^i, j_i)$-cycles, $i = 0, 1, 2, ...$, with multiplier $S < -1$, born from the flip bifurcations of ω_k^j.*

- (b) *Let $(k_1; j_1)$ (embedded representation) be the basic cycle $(S < 1)$ of the rank-one box $\Omega_{k_1}^{j_1}$. For $\lambda \geq \lambda_{k_1}^{*j_1}$, the box $\Omega_{k_1}^{j_1}$ has generated the Cantor set $Cs[k_1; j_1]$, and infinitely many Cantor sets, $Cs[k_1 k_2; j_1, j_2]$, ..., $Cs[k_1 k_2...k_a; j_1, j_2, ...j_a]$, ..., $a = 1, 2, ..., \infty$, from the infinitely many boxes (with fractal organization) embedded into $\Omega_{k_1}^{j_1}$.*

- (c) *For $\lambda \geq \lambda_{k_1}^{*j_1}$, the map T (thus not only $\Omega_{k_1}^{j_1}$) has generated infinitely many Cantor sets related to the infinitely many boxes created for $\lambda \leq \lambda_{k_1}^{*j_1}$. For $\lambda < \lambda_{(k_1)s}^{j_1}$ the map has generated infinitely many Cantor sets related to the infinitely many boxes created for $\lambda \leq \lambda_{(k_1)0}^{j_1}$.*

- (d) *For $\lambda \geq \lambda_1^*$, T has generated all the possible Cantor sets, related to all the basic cycles of T, from the infinitely many boxes embedded into the over all box Ω_1 "*

Denote Λ_λ^*, the fractal set resulting from the union of all the above Cantor sets, and cycles with $S > 1$, $k \geq 3$, including their limit points, generated for $\lambda \geq \lambda_{1s}$. This set satisfies the property $\Lambda_\lambda^* \subseteq \overline{CC_1}$, and its structure is well defined from the knowledge of the Fig. 1 structure. When $\lambda_{1s} < \lambda < \lambda_1^*$, the set Λ_λ^* constitutes the *nucleus of a strange repeller SR*, made up of Λ_λ^* and its increasing rank preimages. From the initial point $x_0 \in \overline{CC_1} \backslash \Lambda_\lambda^*$, n iterations, $n < N$ (N depending on x_0) give rise to a chaotic transient, followed by a regular convergence toward the attractor of T. When $\lambda > \lambda_1^*$, and $x_0 \in]q_1^{-1}, q_1[\backslash \Lambda_\lambda^*$, the set Λ_λ^* generates for $n < N$ a chaotic transient occurring in the interval $]q_1^{-1}, q_1[$, and for $n > N$ the orbit diverges tending toward infinity. On the x-axis, the repelling cycles, and their limits for the period $k \to \infty$, have a fractal organization when $\lambda \geq \lambda_{1s}$. For each point of the parameter λ-axis, $\lambda \geq \lambda_{1s}$, the fractal structure of the map singularities is completely identified from the box-within-a-box bifurcation structure. Consider $\lambda \in \omega_k^j$, with λ sufficiently near $\lambda_{(k)0}^j$ for giving an attracting cycle $(k; j)$. For the map T^k this cycle gives k attracting fixed points P_i, $i = 1, ..., k$, each of them with an immediate basin $D_0(P_i)$, and a total non connected basin $D(P_i) = \cup_{n>0} T^{-n} D_0(P_i)$. The basins $D(P_i)$ are fractal, and have SR as a common limit set.

2.4 General occurrence of the embedded boxes structure

As said in sec. 2.1 the embedded boxes structure generated by the Myrberg's map T (3) also occurs for other types of unimodal maps. Particularly in the case of the general form of quadratic map $y' = ay^2 + 2by + c$ a linear change of variable $y = \alpha x + \beta$ leads to (3) with $\lambda = b^2 - ac - b$, $a\alpha = 1$, $a\beta = -b$. Moreover bimodal maps (with two extrema, i.e. $(Z_1 - Z_3 - Z_1)$ maps) create such bifurcations organization related to each of the two possible attractors (cf. p.401-418 of [6]). For multimodal maps locally this organization exists.

The fractal embedded boxes structure described in secs. 2.1-2.3 shows that if

the Myrberg's map gives a cycle with a period different from 2^i, $i = 0, 1, 2, ...$ (note that $2^2 \neq 4$, $2^3 \neq 8$) then T generated infinitely many repelling cycles and their limit, the whole constituting the nucleus of a strange repeller (see end of sec.2.3). This property occurs for all multimodal maps. It gives a test permitting to conclude to the presence of a strange repeller SR. In the case of a plane map T, if such a cycle exists on a dim 1 manifold, we can say that it contains a strange repeller, generated by the dim 1 map resulting from the restriction of T to this manifold.

3 Singularities and bifurcations specific to plane noninvertible maps

3.1 Critical sets. Chaotic areas

With respect to invertible case the noninvertibility of a map T gives rise to a singularity of a different nature: the *critical set*. The rank-one critical set CS is the geometrical locus of points X having at least two coincident rank-one preimages. Such preimages are located on a set CS_{-1}, the set of coincident rank-one preimages. The set CS satisfies the relations $T^{-1}(CS) \supseteq CS_{-1}$ and $T(CS_{-1}) = CS$. A rank-q critical set CS_{q-1} is given by the rank-q image of CS, $CS_{q-1} = T^q(CS_{-1})$, $CS_0 \equiv CS$. If $\dim X = p = 1$, CS is a rank-one *critical point* C. When $p > 1$ the results obtained till now essentially concern the class of plane maps ($\dim X = p = 2$) [2]. For such maps CS generally becomes a *critical curve* LC. In exceptional cases the critical set may include isolated points. It is the case when the inverse map T^{-1} has a vanishing denominator [11] [12]. The singularity "critical curve" is the fundamental tool for studying two-dimensional noninvertible maps. In general LC is made up of several branches separating the plane into regions Z_i (each of its point has i distinct rank-one preimages). The locus CS_{-1} becomes a curve LC_{-1}, called *rank-one curve of merging preimages*. When the map T is smooth, LC_{-1} is a set of points for which the Jacobian determinant of T vanishes. When LC is made up of only one branch, and separates the plane into two open regions Z_0 and Z_2, we have the simplest case of noninvertible maps, the $(Z_0 - Z_2)$ one (cf. [2] [3] for more complex cases). With respect to the invertible case, the noninvertibility also adds new bifurcation types in a parameter space.

Critical curves permit to define the essential notions of *absorbing area* (d') and *chaotic area* (d) [1]-[3] [6] [9]. Roughly speaking an absorbing area (d') is a region bounded by arcs of critical curves of finite, or infinite, rank LC_n, $n = 0, 1, 2, ...l$, $LC_0 \equiv LC$, such that the successive images of all points belonging to a neighborhood $U(d')$ of (d'), obtained from a finite number of iterations, enter into (d') and cannot get away after entering. Except for some bifurcation cases, a chaotic area (d) is an invariant absorbing area, containing infinitely many sets of repelling cycles of increasing period, and their corresponding limit sets, the lot having a fractal structure. Its boundary ∂d is made up of LC_n arcs. This structure also implies the existence of homoclinic and heteroclinic points (sec. 3.3). Note that a chaotic area may be periodic of period k, i.e. it is constituted by k nonconnected chaotic areas invariant by T^k.

3.2 Basin boundaries

Let A be an attracting set (which can be the union of several attractors), U a neighborhood of A such as $T(U) \subset U$ and $T^n(X) \to A$ when $n \to \infty$, $\forall X \in U$. The open set $D = \bigcup_{n \geq 0} T^{-n}(U)$ is the *total basin* of A. When A is the widest attracting set of a map T, its basin D (called *continent* in [1] [2]) is the open set containing A such that its closure $\overline{D}$ is the locus of points of the plane having bounded trajectories. Its complementary set, denoted by D' (i.e. $\overline{D} \cup D' = \mathbf{R}^2$) when it is non void, is the basin (called *sea* in [1] [2]) of an attracting set at infinity (on the Poincaré's equator), i.e. the locus of points of the plane having divergent trajectories. In such a case the basins D and D' have a common boundary ∂D.

When the map T is invertible, a basin is always *simply connected*. This is not always the case when T is noninvertible, the basin being either *simply connected*, or *multiply connected*, or *nonconnected*, or *both multiply connected and nonconnected*. The nonconnected parts of a basin are called *islands*, and *lakes* are the holes of a multiply connected basin . Consider a $Z_0 - Z_2$ map. Let B be the widest open region of a nonconnected basin containing A, such as $B \cap LC_{-1} \neq \emptyset$, and $\overline{Z}_2$ the closure of Z_2. If B is multiply connected, let ∂B_e be the external boundary of B, and B^* the simply connected region bounded by ∂B_e, i.e. B^* is obtained by removing the holes (lakes) from B. Let $\mathbf{C}(B)$ be the complement of B. The following propositions are proved in [2]:

P1. "(a) B is connected$\Longleftrightarrow B \cap \overline{Z}_2$ is connected $\Longleftrightarrow T(B \cap LC_{-1}) = B \cap LC$. (b) B is nonconnected $\Longleftrightarrow B \cap \overline{Z}_2$ is nonconnected $\Longleftrightarrow T(B \cap LC_{-1}) \subset B \cap LC$, and B is the *immediate basin* D_0 of A. Then $D_0 \cap \overline{Z}_2$ include r areas, called *headlands* Δ_0^r, $r = 1, 2, ...m$, without any connection with the $D_0 \cap \overline{Z}_2$ part containing A. The open domains $D_1^r = T^{-1}(\Delta_0^r)$ are rank-one *islands* (nonconnected with D_0). The union of all the increasing rank preimages of Δ_0^r is the islands set.

P2. "Suppose $B \cap \overline{Z}_2$ is connected. $T^{-1}(B^*) \subset B^* \Longleftrightarrow B$ is multiply connected $\Longleftrightarrow .T(B^* \cap LC_{-1}) \supset B^* \cap LC$. Then each dim 1 largest connected component of $T(B^* \cap LC_{-1}) \backslash (B^* \cap LC)$ belongs to a dim 2 close set belonging to $\mathbf{C}(B) \cap \overline{Z}_2$, called a *bay* denoted H_0^r, $r = 1, 2, ...m$. The rank-one preimages of H_0^r, $H_1^r = T^{-1}(H_0^r)$, $H_0^r \cap LC_{-1} \neq \emptyset$, are the rank-one *lakes* of the multiply connected basin D. The union of the increasing rank preimages of H_0^r is the lakes set.

The corresponding qualitative figures can be found in [1] [2]. Maps of $(Z_1 - Z_3 - Z_1)$ type give equivalent propositions changing Z_0 into Z_1, Z_2 into Z_3, and $\Longleftrightarrow$ into $\Rightarrow$. Then a basin may be also nonconnected (resp. multiply connected) without creation of a headland (resp. bay). Let L and L' be the two branches of LC, $LC = L \cup L'$, bounding the region Z_3. If $B \cap L \neq \emptyset$ and $B \cap L' = \emptyset$ then B is nonconnected. If $B \cap L \neq \emptyset$ and $B \cap L' \neq \emptyset$ then B is connected. More details are given in [1] [2] [13]. In the case of $(Z_0 - Z_2)$ and $(Z_1 - Z_3 - Z_1)$ maps, global bifurcations related to *basin transitions* "simply connected $\longleftrightarrow$ nonconnected, or "simply connected $\longleftrightarrow$ multiply connected, or "multiply connected $\longleftrightarrow$ nonconnected" are described in the above references. Whatever be the map, invertible (with $p \geq 2$), or noninvertible (with $p \geq 1$), a basin boundary $\partial D(A)$ may be also fractal, i.e. it has a non integer dimension, $\partial D(A)$ containing a strange repeller.

3.3   Homoclinic and heteroclinic situations

Let U be a neighborhood of an unstable fixed point (saddle, node or focus) p^*. The *local* (i.e. belonging to U) *unstable set* of p^* is defined as the locus of points of U having a subset of increasing rank preimages which tend toward p^*. The *global unstable set* is the locus of all the points having a sequence of preimages which converges toward p^*. It is invariant and obtained by constructing the images of the local unstable set. When the map T is continuous and noninvertible, the unstable set $W^u(p^*)$ of a saddle point p^* is connected and self intersections may occur (so that it may not be a manifold), which cannot happen for invertible maps. Self intersections and loops of $W^u(p^*)$ are described in [2] (p. 506-515), [6] (p. 373-374), [8] (p. 203-222). The role of critical curves is again essential for understanding the formation of self intersections of the unstable set of a saddle, and also the properties of invariant closed curves.

   The *stable set* $W^s(p^*)$ of a saddle p^* is backward invariant $T^{-1}[W^s(p^*)] = W^s(p^*)$. It is mapped onto itself by T, $T[W^s(p^*)] \subseteq W^s(p^*)$. The stable set is invariant if T is invertible, while for a noninvertible map it may be strictly mapped onto itself. When T is continuous, $W^s(p^*)$ self intersections cannot occur. Then $W^s(p^*)$ can be called manifold, being a connected manifold, or the union of disjoint connected components which are manifolds. While in invertible maps an expanding fixed point (unstable node, or focus) p^* has no stable set, when T is noninvertible the stable set of p^* is given by all the preimages (of any rank) of this point: $W^s(p^*) = \bigcup_{n \geq 0} T^{-n}(p^*)$ [2]. So if T is noninvertible $W^s(p^*)$ *may be non-connected and made up of infinitely many closed curves*, passing through the increasing rank preimages of p^*. A point q is said *homoclinic* to the non-attracting fixed point p^* (or homoclinic point of p^*) iff $q \in W^s(p^*) \cap W^u(p^*)$. *Heteroclinic points* are obtained when the stable and unstable sets are related to two different fixed points. A "*contact bifurcation*" (between a critical set and a stable (or unstable) set) may correspond to homoclinic and heteroclinic bifurcations, in this framework critical curves are useful for interpreting such problems, and the related bifurcations [2]. Homoclinic and heteroclinic sets being fractal are source of strange repellers.

4   Plane maps with type A strange repeller

4.1   Generalities

A strange repeller SR_A of *type A* is defined as generated by a m-modal dim 1 map resulting from the restriction T_r of a plane map T to a manifold arc. This means that the SR_A fractal dimension d_f is $0 < d_f < 1$. The simplest examples are those for which one of the two axes is invariant. It is the case of the two following maps T:

$$x' = ax(1 - 2x - y),\ y' = bxy \tag{4}$$

$$x' = x^2 + y - \lambda,\ y' = \delta x + \gamma y + 4x^2 y \tag{5}$$

   With the first map the x-axis is invariant, the restriction T_r of T to $y = 0$ being $x' = ax(1 - 2x)$. The second map with $\delta = 0$ [14] also gives the x-axis
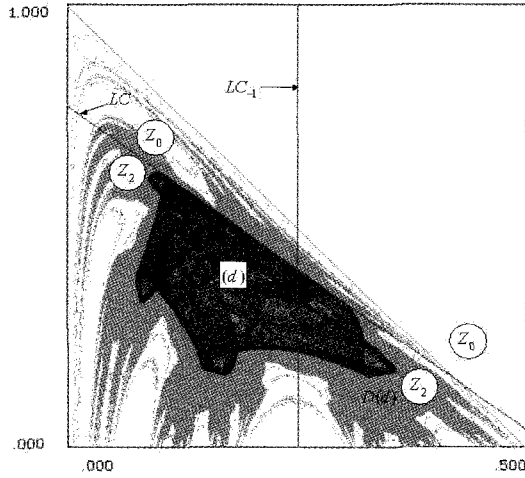
Figure 2. Map(4). $a = 4.2$, $b = 6.16$. Basin $D(d)$ of the chaotic area (d).

invariant, the restriction T_r to $y = 0$ being the Myrberg's map $x' = x^2 - \lambda$. The two maps generate the box-within-a-box fractal bifurcation structure of Fig. 1, taking into account $a = 1 + \sqrt{1 + 4\lambda}$ (cf. sec. 2.4), $a > 0$. Each of the two maps T_r generates a strange repeller SR_A as soon as a, or λ, exceeds the limit of flip bifurcation values belonging to the first Myrberg' *spectrum* ω_1 (cf. sec. 2.1). This value is $\lambda = \lambda_{1s} \simeq 1.401155189$, for $x' = x^2 - \lambda$, $a = a_{1s} \simeq 3.569945929$ for $x' = ax(1 - 2x)$. When $\lambda \geq \lambda_{1s}$ and $a \geq a_{1s}$ infinitely many repelling cycles and their limits, constituting the nucleus of SR_A, are created on the x-axis (cf. sec. 2.3).

4.2 "Bordering fractals" situations

Figure 2 represents the basin $D(d)$ of a chaotic area (d) generated by the $(Z_0 - Z_2)$ map (4) for $a = 4.2$, $b = 6.16$. The a-value of the map T_r is such that $a > a_1^* = 4$ equivalent to the situation $\lambda > \lambda_1^*$ defined in sec. 2. So the strange repeller SR_A is the *completely disjointed* Julia-Fatou set (i.e. a Cantor set) located inside the segment $[O, O^{-1}]$, $O(0;0)$, $O^{-1}(1/2;0)$ (cf. sec. 2.1). The fixed point O is a saddle (multipliers $S_1 = 0$, $S_2 = a$). The basin $D(d)$ is nonconnected. It is not fractal (because without self similarity properties), but its boundary $\partial D(d)$ is fractal because $\partial D(d) \cap (y = 0)$ is a Cantor set. A more complete study of another quadratic form of this map is given in [15]. The cubic map $x' = \alpha x(1 - 2x^2 - y)$, $y' = \beta xy$ has equivalent properties (see [16] for $\beta = 3.3$), but with a strange repeller generated by a symmetric cubic bimodal map ($Z_1 - Z_3 - Z_1$ type studied, with its box-within-a-box fractal structure, p. 415-418 of [6]). We call "*bordering fractal*" such a basin boundary $\partial D(d)$.

Consider the following quadratic $(Z_0 - Z_2)$ map T (cf. a first study in [17])
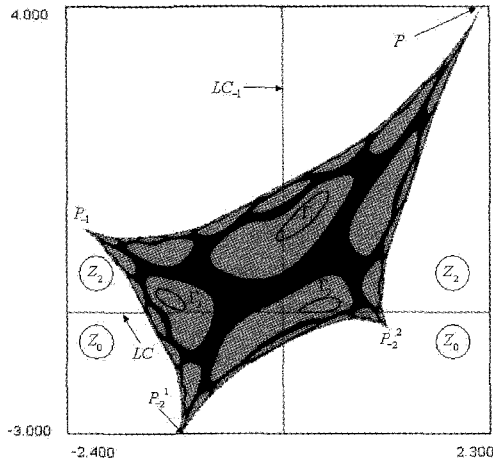
$$x' = ax + y, \ y' = b + x^2 \tag{6}$$

Figure 3. Map(6). $a = -0.8$, $b = -1$. Basin (dark grey) of the fixed point Q, and basin (light grey) of the period 3 invariant closed curve Γ_i, $i = 1, 2, 3$.

It is particularly rich in examples leading to fractal situations [2]. The curve LC_{-1} where two rank-one preimages merge is $x = 0$, and the critical curve is the line $y = b$.

A fold bifurcation occurs when $a = -0.8$, $b = b_f \simeq -0.87001$. For $b = b_f + \varepsilon$, $\varepsilon > 0$ sufficiently small, only one attractor exists: a stable fixed point Q. Its basin boundary $\partial D(Q)$ contains cycles not having the period 2^i, which indicates the existence of a strange repeller SR_A on this boundary generated by the dim 1 noninvertible map restricted to $\partial D(Q)$ (cf. sec. 2.4). For example let $a = -0.8$, $b = -0.867$ be the parameters values. The point P ($x \simeq 2.195$, $y \simeq 3.950$) is an unstable node ($S_1 = 1.733$, $S_2 = -2.533$), cusp for the basin boundary $\partial D(Q)$. The points P_{-1}, $P_{-2}^1 \cup P_{-2}^2$ are respectively the ranks one and two preimages of P, and $T(\overline{PP_{-2}^2}) = T(\overline{PP_{-1}})$. By application of T^2, the map T_r^2 restricted to the arc $\overline{PP_{-1}}$ is unimodal, its extremum being the point $LC_{-1} \cap \partial D(Q)$. According to sec. 2.4 the presence of two period 20 unstable nodes, i.e. period 10 cycles for T^2 and T_r^2 (so not having the period 2^i), ensures the existence of a strange repeller SR_A on the arcs $\overline{PP_{-1}}$, $\overline{PP_{-2}^2}$. One point of the first $k = 20$ cycle is $x \simeq 0.504$, $y \simeq 1.462$ with multipliers $S_1 = 29302$, $S_2 = 13.34$. For the second cycle one has $x \simeq 0.267$, $y \simeq 1.667$ with multipliers $S_1 = 28593$, $S_2 = -10.57$.

The parameter value $b = b_f$ is a fold bifurcation, giving rise for $b = b_f - \varepsilon$ to a pair of period three cycles (saddle and stable node) coexisting with the stable fixed point Q, inside the previous domain bounded by $\partial D(Q)$. With decreasing values of b the period 3 stable node turns into a stable period 3 focus, the destabilization of which leads to a stable period 3 invariant closed curve Γ_i, $i = 1, 2, 3$ via a Neimark's bifurcation. Let us denote by A the attracting set of T, made up of the union of Q and the stable period 3 set U^3 (cycle, or Γ_i). Let $D(A)$ be the A basin, $\partial D(A)$ its boundary, $D(A) \cup \partial D(A)$ being the bounded orbits set. The

immediate basin $D_0(U^3)$ of U^3 is nonconnected inside $D(A)$, and so the noncon-nection of the total basin $D(U^3) \subset D(A)$ does not result from the creation of a headland. The basin $D(Q)$ of Q is simply connected as the basin $D(A)$. Figure 3 $(a = -0.8, b = -1)$ shows the two basins $D(Q)$ and $D(\Gamma_i)$ when $U^3 \equiv \bigcup_{i=1}^{3}(\Gamma_i)$. The new situation is characterized by the fact that the limit set of the basin $D(U^3)$, $\lim_{n\to\infty} T^{-n}[D_0(U^3)]$, is the strange repeller nucleus inside $\overline{PP_{-1}} \cup \overline{PP_{-2}^2}$ and its in-creasing rank preimages located on $\partial D(A)$, the whole constituting SR_A. The basin $D(Q)$ has a denditric aspect when $D_0(U^3)$ is sufficiently large. Such a situation of the basins $D(Q)$ and $D(U^3)$ is also of "bordering fractal" type.

5 Fractal boundary with type C strange repeller

A strange repeller SR_C of *type C* is defined as generated from the destruction of a chaotic area (d) (cf. sec. 3.1). This would mean that the SR_C fractal dimension d_f is $1 < d_f < 2$. In general they coexist with type-A ones located on the external boundary of a basin. An example is given by the map T (6) with a fixed value $a = -0.42$, the parameter b decreasing from -1.08. For $b = -1.08$ the map T has two attractors: an invariant closed curve Γ and a stable period three cycle Q_i, $i = 1, 2, 3$. their basins boundary being bordering fractal due to a type-A strange repeller SR_A, with a nucleus N^{rA}, located on the external boundary of the union of the two basins. When b decreases the period three cycle becomes unstable giving rise to a stable period three invariant closed curve Γ_i, which after turns into a period three chaotic area (d_i), $i = 1, 2, 3$ (see Fig. 4 $a = -0.42$, $b = -1.09$). For $b = b_c$, $-1.1182 > b_c > -1.1183$ a contact bifurcation [2], between the boundary $\partial(d_i)$ of the chaotic area (d_i) and the boundary of its immediate basin, destroys the chaotic area as an attracting set. When $b = b_c - \varepsilon$, $\varepsilon > 0$ sufficiently small, the (d_i) destabilization gives rise to a strange repeller SR_C, the nucleus N_i^{rC} $(i = 1, 2, 3)$ of which contains infinitely many sequences of unstable cycles with increasing period, and their limit sets (cf. Fig. 4b, $a = -0.42$, $b = -1.25$). The repeller SR_C is the union of N_i^{rC} and the increasing rank preimages of N_i^{rC}. The map T has a unique attractor: a stable period seven cycle F_j, $j = 1, ..., 7$, located around the unstable focus Q, with a simply connected basin $D(F_j)$. Even if this does not appear by a numerical scan of the plane, *the basin boundary $\partial D(F_j)$ is fractal* as made up of an "external" closed part $\partial_e D(F_j)$ union of four arcs $\overline{PP_{-1}}$, $\overline{PP_{-2}^2}$, $\overline{P_{-1}P_{-2}^1}$, $\overline{P_{-2}^1 P_{-2}^2}$ (Fig. 4), and also of SR_C located inside the open domain bounded by $\partial_e D(F_j)$, i.e. $\partial D(F_j) = \partial_e D(F_j) \cup SR_C$. The external boundary $\partial_e D(F_j)$ contains the strange repeller SR_A $(0 < d_f < 1)$ due to the presence of cycles not having the period 2^i.

When b decreases a global bifurcation occurs: *the simply connected basin $D(F_j)$ turns into a multiply connected one, after the critical curve LC becomes tangent to $\partial_e D(F_j)$*, i.e. now $\partial_e D(F_j) \cap LC \neq \varnothing$. Lakes (holes) of $D(F_j)$ are created with a fractal structure (see Fig. 5 and the Fig. 6 enlargements), their limit set being constituted by the two strange repellers $SR_A \cup SR_C$. These lakes are the increasing rank preimages $H_n^{i_1 \cdots i_{n-1}}$, $i_r = 1, 2$, $n = 1, 2, 3, ..$, of the bay H_0. Moreover the arborescent sequence of the increasing rank preimages of the SR_A cycles (belonging to $\partial_e D(F_j)$) now tends toward $SR_A \cup SR_C$. Continuing the b decrease a new bifurcation takes place when the cusp point $P_{-2}^2 \in T^{-2}(P)$ has a
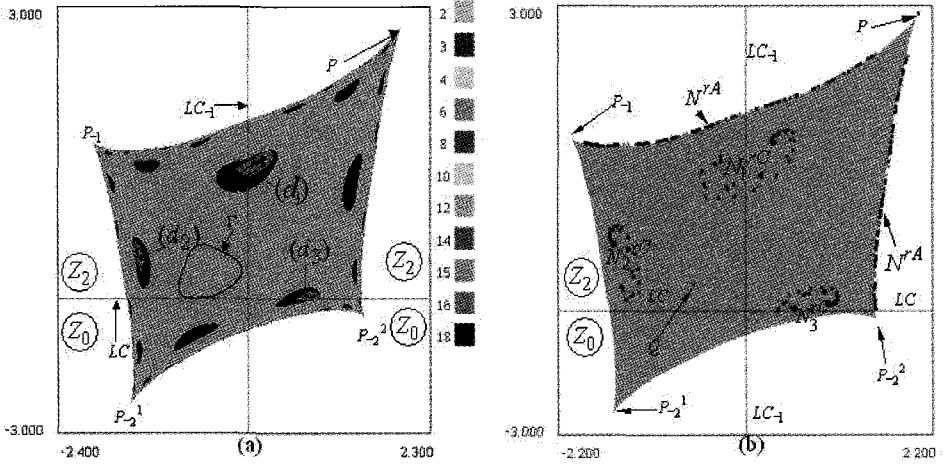
Figure 4. Map(6).(a) $a = -0.42$, $b = -1.09$, chaotic area (d_i) (light grey), $i = 1, 2, 3$, and its basin (black), Γ basin (grey). (b) $a = -0.42$, $b = -1.25$, (d) has been destroyed. Period-k cycles,$1 < k < 19$, belonging to the nucleus N_i^{rC} (resp. N^{rA}) of the strange repeller SR_C (resp. SR_A). SR_C makes fractal the basin boundary. Left column of grey shades gives k.

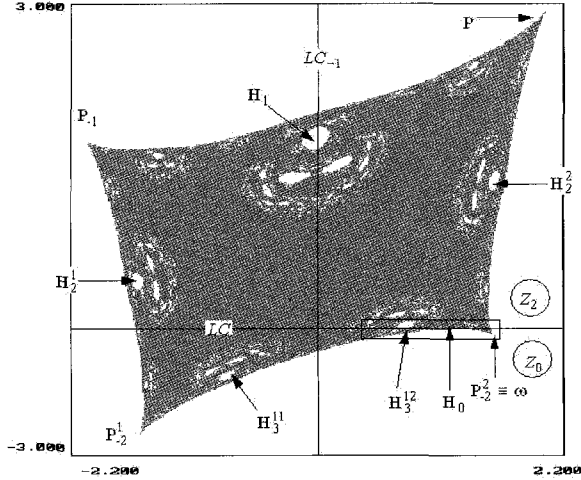

Figure 5. Map(6).$a = -0.42$, $b = -1.31$. Multiply connected basin. The lakes (basin holes) limit set is the strange repellers union $SR_C \cup SR_A$.

contact with LC, which after $(P_2^{-2} \in Z_2)$ leads to an opening of the *lakes* in the *sea* (domain of divergent orbits), called *roadsteads* [1] [2]. First $D(F_j)$ is non connected with small islands inside the roadsteads, and the boundary $\partial_e D(F_j) \equiv \partial D(F_j)$ is fractal, nowhere smooth (in Figs 5-6 $\partial_e D(F_j)$ is piecewise smooth). With a new b decrease $D(F_j)$ becomes simply connected, $\partial D(F_j)$ remaining fractal and
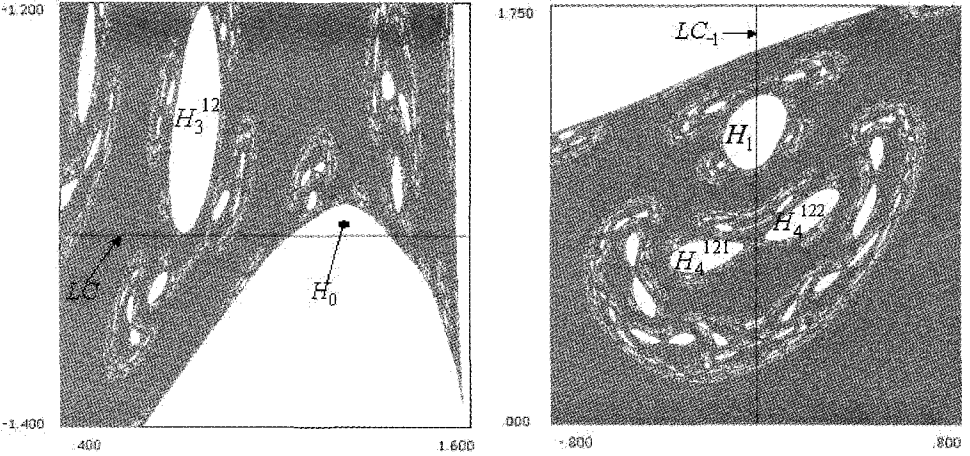
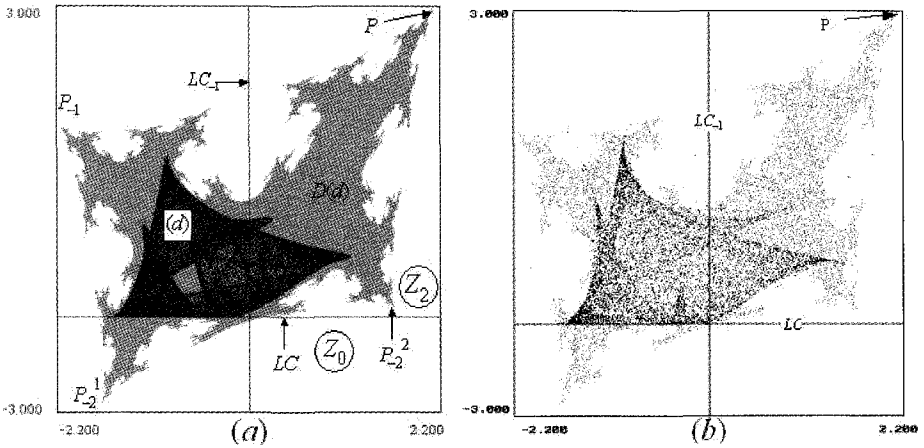Figure 6. Map(6). $a = -0.42$, $b = -1.31$. Enlargements of Fig.6.

Figure 7. Map(6). **(a)** $a = -0.42$, $b = -1.6$, basin $D(d)$ of the chaotic area (d) . **(b)** $a = -0.42$, $b = -1.65$, (d) is destroyed, chaotic transient before divergence.

the map attractor undergoes a sequence of bifurcations (described in chapter 6 of [2]) which leads to a chaotic area (d) (Fig. 7a $a = -0.42$, $b = -1.6$). The value $b = b_f \simeq -1.63892$ gives rise to a contact between the basin boundary $\partial D(d)$ of (d) and the LC segment belonging to the boundary $\partial(d)$ of the chaotic area (d). For $b < b_f$ the attractor (d) is destroyed, creating a new strange repeller inside the domain of divergent orbits. It results from the union of SR_A, SR_C and the strange repeller issued from the destruction of (d). Depending on the initial condition this new strange repeller defines a double region of chaotic transient (occupying the previous (d) and $D(d)$ areas), before a regular divergence of the orbits (cf. Fig. 7b,

$a = -0.42$, $b = -1.65$).

6   Class related to the complex dim 1 map and to the derived ones

A complex one-dimensional map can be written in the form of a two-dimensional real one T_Z. Then the left parts of the two relations defining T_Z verify the Cauchy-Riemann conditions. Derived maps from T_Z are those obtained by a small perturbation of T_Z, the Cauchy-Riemann conditions being no longer satisfied. This means that the strange repeller SR fractal dimension d_f is $1 < d_f < 2$. Such strange repellers, denoted SR_B, will be said of "*type B*". For example the quadratic complex map Z: $z' = z^2 - \lambda$ (λ real parameter) $z = x + jy$, $j^2 = -1$, is also the plane noninvertible map T_Z: $x' = x^2 - y^2 - \lambda$, $y' = 2xy$. The basin boundary ∂D of the attractor generated by T_Z is the well known Julia-Fatou fractal perfect set (E) (nowhere smooth) made up of all the repelling cycles, their limit sets, and their increasing rank preimages. When $\lambda \geq 2$ this fractal set is located on the interval $-2 \leq x \leq 2$ of the x-axis with $0 < d_f \leq 1$, i.e. all the cycles generated by the complex map $z' = z^2 - \lambda$ and their limit sets are real. These cycles are those of the Myrberg's map $x' = x^2 - \lambda$, identified by the symbolism $(k; j)$ as coming from the fractal box-within-a-box bifurcation structure described in sec. 2, .

The map T (5) gives rise to regions Z_i, $i = 0, 2, 4$ (see Figs. 8-10). For $\delta = \gamma = 0$ the basin boundary structure generated by T is particular and well known in the invariant half-plane $\Pi_- = \{(x, y) : y < 0\}$ [14]. Indeed there exists a semiconjugacy between the map T and the complex quadratic map Z, defined by $z' = z^2 - \lambda$. Then inside Π_- T is equivalent to the real two-dimensional map T_Z. For $\lambda = 1$, $\gamma = \delta = 0$ the map T has two attractors: a fixed point P in the region $y > 0$, and a period two cycles $a_i = 1, 2$ on the x-axis (the one of the Myrberg's map). Their basins $D(P)$ and $D(a_i)$ are light grey and dark grey colored in Fig. 8a (symmetric with respect to $x = 0$), which shows that the basin boundary $\partial D(a_i) \cap \Pi_-$ is fractal reproducing the Julia-Fatou perfect set (nowhere smooth) (E) of the map $z' = z^2 - 1$. The boundaries $\partial D(a_i)$ and $\partial D(P)$ inside the half-plane $y > 0$ are smooth. Figure 8b,c shows how $\partial D(a_i) \cap \Pi_-$ is perturbed when γ has a sufficiently small positive value ($\lambda = 1$, $\gamma = 0.005$, $\delta = 0$). Due to the formation of the headland Δ_0 infinitely many islands are created, the first one being $D_1 = T^{-1}(\Delta_0)$. These islands fractally recur. Their limit set is a strange repeller SR issued from the set (E) which has lost sequences of Cantor sets $Cs[k, j]$ (cf. sec. 2.3) whose cycles have turned into complex ones. Figures 9a,b and 10 ($\lambda = 1$, $\gamma = 0.00005$, $\delta = 0.009$) give another behavior of the map (5). In addition to the headland Δ_0 with its related islands, the map creates two bays H_0 and H_0' with their related lakes. The islands inside H_0 and H_0' are injected in the lakes (H_0' having ordinates higher than 1.5 does not appear in the figure), the nuclei of the related strange repellers being located in $\partial D(a_i) \cap H_0$ and $\partial D(a_i) \cap H_0'$.

7   Conclusion

This text gives only a succinct view of the strange repellers (fractal set) role in the fractalization of basin boundaries. Section 2.3 has shown this role for a one-
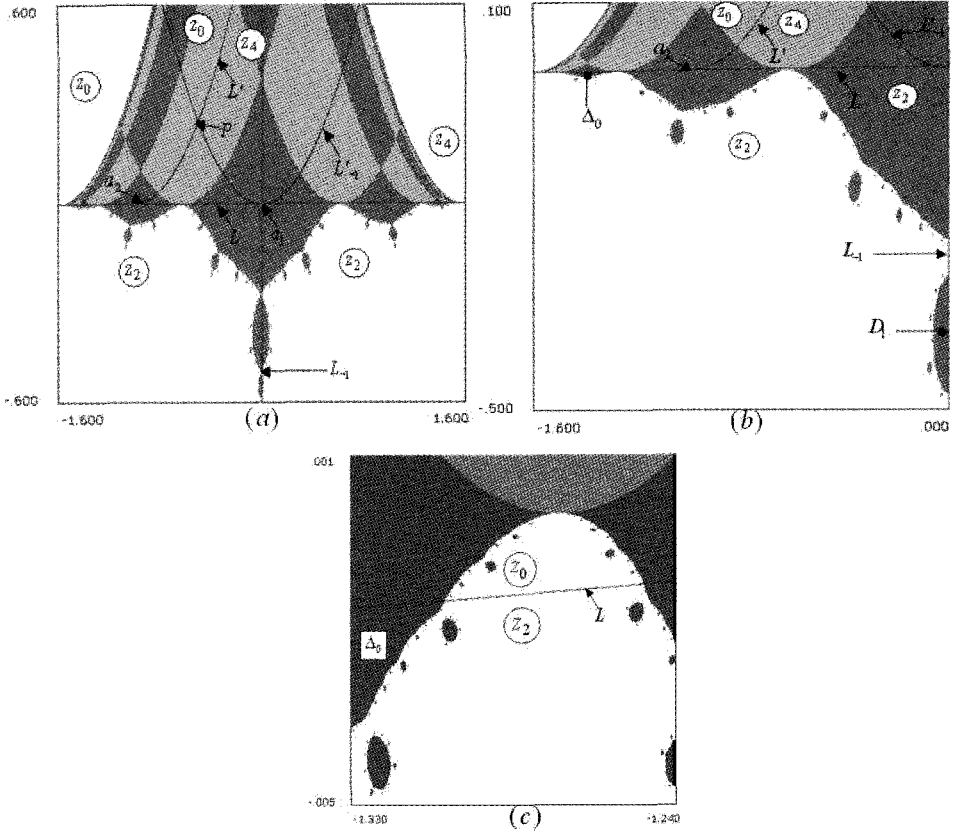
Figure 8. Map (5). (a) $\lambda = 1$, $\gamma = \delta = 0$. $\partial D(a_i) \cap (y < 0)$, $i = 1, 2$, is a Julia-Fatou set. (b) $\lambda = 1$, $\gamma = 0.005$, $\delta = 0$, $D(a_i)$ is non connected. (c) Enlargement of (b) showing the headland Δ_0.

dimensional map. For plane maps T the simplest cases correspond to only one repeller denoted SR_A (with a dimension $0 < d_f < 1$) located on a *smooth* arc of the boundary of a basin D. In this situation SR_A is created by a dim 1 multimodal map related to the restriction of T to this arc, which gives rise to the fractal bifurcation structure described in sec. 2. Strange repellers of type B and C, denoted SR_B and SR_C, have a dimension $1 < d_f < 2$. The paper gives examples of such fractal repellers, SR_B resulting from a perturbation of a dim 1 map with complex variable, SR_C resulting from the destruction of a chaotic area. In this last case as soon as SR_C appears (so out of a smooth arc belonging to the "external" boundary of D) it becomes a subset of the complete basin boundary ∂D, which is fractal even if a numerical simulation by plane scanning cannot show this situation. The ∂D fractalization appears numerically after a global bifurcation due to the contact of ∂D with the critical curve LC. Then SR_C belongs to the limit set of lakes (holes) of the basin if is multiply connected, or islands (nonconnected parts) if it is
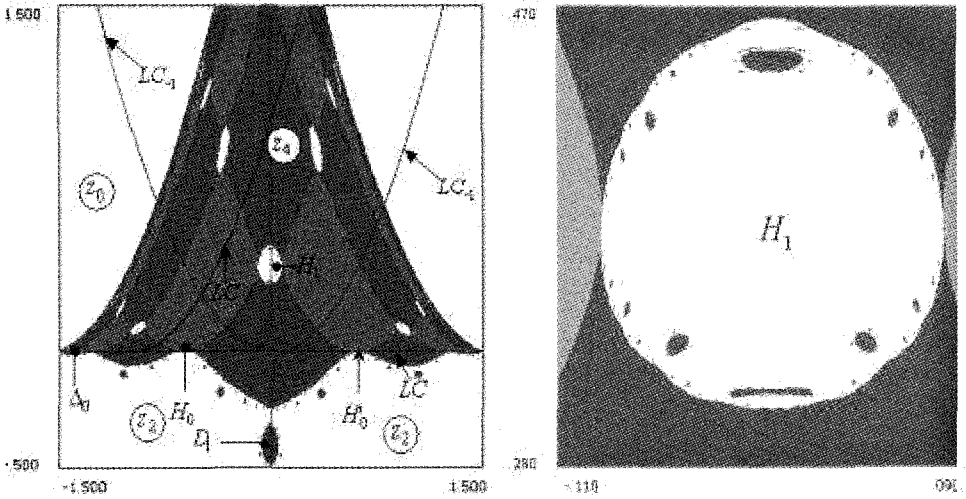
Figure 9. Map (5), $\lambda = 1$, $\gamma = 0.00005$, $\delta = 0.009$, basins $D(P)$ and $D(a_i)$. $D(a_i)$ is nonconnected due to the headland Δ_0 and multiply connected due to the two bays H_0, H_0'. The enlargement on the right shows the rank-one lake H_1, and islands inside H_1.
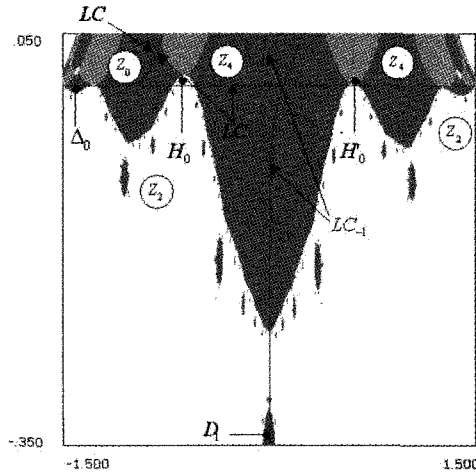
Figure 10. Map (5), $\lambda = 1$, $\gamma = 0.00005$, $\delta = 0.009$. Other enlargement showing the two bays H_0, H_0'.

nonconnected. A lot of different situations, not described in this text, occur. They differ by the fact that several strange repellers of different types may coexist, by the nature of ∂D, and by the location of LC with respect to ∂D. Some of these situations are shown in [1] [2] [13].

Acknowledgments

References

1. C. Mira, D. Fournier-Prunaret, L. Gardini, H. Kawakami, & J.C. Cathala. *Int. J. of Bif. and Chaos*, 4(2), 343-381,(1994).

2. C. Mira, L. Gardini, A. Barugola, & J.C. Cathala. *Chaotic dynamics in two-dimensional noninvertible maps.* World Scientific Series on Nonlinear Sciences, Series A, vol. 20, 630 pages, (1996).

3. C. Mira, G. Millerioux, J.P. Carcasses, & L. Gardini. *Int. J. of Bif. and Chaos*, 6(8), 1439-1462, (1996).

4. C. Mira, *Proceedings of the 7th International Conference On Nonlinear Oscillations*, Berlin 1975. Akademic Verlag, Berlin 1977, Band I 2, 81-93, (1975).

5. I. Gumowski, & C. Mira. *Comptes Rendus Acad. Sc. Paris,.* **A281**, 45-48, (1975).

6. I. Gumowski, & C. Mira. *Dynamique chaotique. Transformations ponctuelles. Transition ordre désordre.* Cépadues Editions Toulouse, 480 pages, (1980).

7. J. Guckenheimer. *N.Y. Acad. of Sci.*, **75**(1), 343-347, (1980).

8. I. Gumowski, & C. Mira. *Recurrences and discrete dynamic systems.* Lecture Notes in Mathematics, 809, Springer Verlag, 250 pages, (1980).

9. C. Mira. *"Chaotic dynamics. From the one-dimensional endomorphism to the two-dimensional diffeomorphism".* World Scientific, Singapore, 450 pages, (1987).

10. P.J. Myrberg. *Ann. Acad. Sci.. Fenn.*, **A**(336), 1-10,(1963).

11. G.I. Bischi, L. Gardini, & C. Mira. *Int. J. of Bif. and Chaos*, **9**(1), 119-153 (1999).

12. G.I. Bischi, L. Gardini, & C. Mira. *Int. J. of Bif. and Chaos*, **13**(8), 2253-2277, (2003).

13. G.I. Bischi, L. Gardini, & C. Mira. "Basin fractalizations generated by a two-dimensional family of $(Z_1 - Z_3 - Z_1)$ maps". *Int. J. of Bif. and Chaos*, to appear (2006).

14. A. Agliari, L. Gardini & C. Mira. *Int. J. of Bif. and Chaos*, **13**(7), 1767-1785, (2003).

15. Gu E.-G. "Global bifurcations of domains of feasible trajectories" *Int. J. of Bif. and Chaos*, to appear (2006).

16. J.C. Cathala. *Int. J. of Bif. and Chaos*, **13**(4),905-933, (2003).

17. H. Kawakami & K. Kobayashi. *Bull. Fac. Engin. Tokushima University*, **16**, 29-46 (1979).

A GENERATIVE CONSTRUCTION AND VISUALIZATION OF 3D FRACTAL MEASURES

TOMEK MARTYN

Computer Graphics Laboratory, Institute of Computer Science,
Warsaw University of Technology, ul Nowowiejska 15/19,
00-665 Warsaw, Poland
E-mail: martyn@ii.pw.edu.pl

An approach to volume rendering 3D fractal measures is presented. Visualization is done with the adaptive level of detail at pixel-size accuracy. Using the IFS representation we show a generic construction of the measures, which can be thought as a variant of Carathéodory's method when applied to the fractal context. Our construction holds a wide range of fractal measures, including the RIFS invariant measures. Then, we introduce the measure instancing technique, which allows one to approximate the measure of a set at any accuracy required. The technique enables fractal measures to be ray-traced with low memory and without storing a volume of cells in memory resources. Instead, we utilize a dynamic hierarchy of balls on the object instancing principles. Rather than with semi-lines, the measure is ray-traced with cones so as to speed up computation and to prevent aliasing. To illuminate measures, we use a generalization of the volume rendering integral which we adapt to the case of measures.

1. Introduction

In this paper we investigate the problem of direct volume rendering of measures supported by fractals specified by 3D affine IFSs. We present a generic construction of measures on IFS fractals, which can be considered a fractal variant of the general construction by Carathéodory[2]. Our method resembles also the repeated subdivision method[3] when applied to the IFS context. Then, we show that the class of the IFS and RIFS invariant measures, is just the special case of our general recipe. We also define another class of fractal measures which arise from our construction and the Perlin noise function[5]. Next, we introduce measure instancing which can be regarded as object instancing[6] used in the measure-theoretic context. The technique approximates at any accuracy the value of the measure of a given subset. In this work we use it to approximate, at pixel-size accuracy, the value of our version of the volume rendering integral that we adapt to illuminate measures. Measure instancing allows a measure to be rendered volumetrically without any volume divided into cells that has to be stored in memory. Instead, the role of a volume is played by a dynamical tree of balls, whose single paths are created and removed during rendering, on similar principles when "solid" IFS attractors are ray-traced[6-7]. This allows for the adaptive level of detail at pixel-size accuracy.

103

2. Preliminaries

2.1. *Iterated function systems*

A map $w\colon X \to X$ on a metric space (X, d) is *Lipschitz continuous* if there is $\lambda \in [0, \infty)$ such that $d(w(x), w(y)) \leq \lambda d(x, y), \forall x, y \in X$. The smallest λ for which the inequality holds is called the *Lipschitz constant* of w; we denote it by $\lambda(w)$. If $\lambda(w) < 1$ then w is called a *contractive map* (or a *contraction*), and $\lambda(w)$ itself is termed the *contractivity factor* of w. If (X, d) is complete then, by the fixed point theorem, each contraction possesses exactly one fixed point, i.e. $x = w(x)$, $x \in X$.

An *iterated function system* (IFS) consists of a finite number of contractions on a complete metric space (X, d) and it is usually written as $(X; \{w_i\}_{i=1}^N)$. The *Hutchinson operator* associated with the IFS is the map which acts on the space $\mathcal{H}(X)$ of all compact and nonempty subsets of X, and it is defined by $W(E) := \bigcup_{i=1}^N w_i(E)$. One can show that $\mathcal{H}(X)$ equipped with the Hausdorff metric is a complete metric space, and W is a contraction on it[1]. Hence, by the fixed point theorem, there is a unique nonempty compact set $A \subset X$ which is invariant under W, i.e. $A = \bigcup_{i=1}^N w_i(A)$. The set is called the *attractor* of the IFS $(X; \{w_i\}_{i=1}^N)$.

2.2. *Measures*

A *measure* μ on a space X is a set function $\mu\colon 2^X \to [0, \infty]$ which satisfies the three following conditions: 1) $\mu(\emptyset) = 0$; 2) $\mu(A) \leq \mu(B)$ whenever $A \subset B \subset X$; 3) $\mu(\bigcup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty \mu(A_i)$ whenever $A_i \subset X$. Very often it is convenient to think of a measure as a function which specifies a distribution of a mass on space.

A set $S \subset X$ is *μ-measurable* if $\mu(E) = \mu(E \cap S) + \mu(E \setminus S)$, for any $E \subset X$. A measure μ on a metric space (X, d) is a *Borel* measure if all Borel subsets of X are *μ-measurable*.

The *support* of a measure μ on a metric space X is defined by $\operatorname{spt} \mu := X \setminus \bigcup \{V : V \text{ is open and } \mu(V) = 0\}$. If μ is a Borel measure and X is separable (examples are $\mathbb{R}^n$ and any compact space) then $\mu(X \setminus \operatorname{spt} \mu) = 0$ [2]. Thus, in such a case

$$\mu(E) = \mu(E \cap \operatorname{spt} \mu), \quad \forall E \subset X \tag{1}$$

because $\operatorname{spt} \mu$ is closed and hence Borel, and $\mu(E \setminus \operatorname{spt} \mu) = 0$, for $E \setminus \operatorname{spt} \mu \subset X \setminus \operatorname{spt} \mu$.

3. Constructing measures on IFS fractals

Let $\mathcal{A}$ denote the *address space*[1] on N symbols, that is $\mathcal{A}$ consists of semi-infinite words $\sigma = \sigma_1 \sigma_2 \ldots \sigma_m \ldots$, where $\sigma_i \in \{1, \ldots, N\}$. By $\sigma|n$ we denote the prefix of σ of length n, i.e. $\sigma|n = \sigma_1 \sigma_2 \ldots \sigma_n$, $n \in \mathbb{N}$, where by convention $\sigma|0$ represents the empty word. Now, given an IFS $(X; \{w_i\}_{i=1}^N)$ with the attractor A, we define a function by

$$\phi(\sigma|n) := \begin{cases} w_{\sigma_1} \circ \ldots \circ w_{\sigma_n} & \text{if } n > 0 \\ \text{identity map} & \text{otherwise} \end{cases},$$

for every $\sigma \in \mathcal{A}$ and $n \in \mathbb{N}$.

Let $\mathcal{F} = \{A_{\sigma|n}\}$, $\sigma \in \mathcal{A}$, $n \in \mathbb{N}$, be the family of sets specified by the attractor subsets $A_{\sigma|n} \equiv \phi(\sigma|n)(A)$. Now, let $\{\mu_{\sigma|n}\}$, $\sigma \in \mathcal{A}$, $n \in \mathbb{N}$, be a collection of finite Borel measures on X such that for every $n \in \mathbb{N}$ and $\sigma \in \mathcal{A}$, the following two conditions hold: 1) $\operatorname{spt}\mu_{\sigma|n} \subset A_{\sigma|n}$, and 2) $\mu_{\sigma|n} = \sum_{\sigma_{n+1}=1}^{N} \mu_{(\sigma|n)\sigma_{n+1}}$, where $(\sigma|n)\sigma_{n+1} = \sigma_1 \ldots \sigma_n \sigma_{n+1}$.

By definition, $\mu \equiv \mu_{\sigma|0}$ is a finite Borel measure on X, which is supported by a subset of the IFS attractor. Moreover, μ can be decomposed into a finite sum of Borel measures with supports of diameters not greater than $\varepsilon > 0$. The latter is due to the fact that IFS attractors are self-tiled sets, which using the notation of this paper can be written as

$$
\begin{aligned}
A_{\sigma|n} = \phi(\sigma|n)(A) = \phi(\sigma|n)\Big(\bigcup_{\sigma_{n+1}=1}^{N} w_{\sigma_{n+1}}(A) \Big) &= \bigcup_{\sigma_{n+1}=1}^{N} \phi(\sigma|n) \circ w_{\sigma_{n+1}}(A) \\
&= \bigcup_{\sigma_{n+1}=1}^{N} \phi\big((\sigma|n)\sigma_{n+1}\big)(A) = \bigcup_{\sigma_{n+1}=1}^{N} A_{(\sigma|n)\sigma_{n+1}}
\end{aligned}
\tag{2a}
$$

$$
\operatorname{diam}(A_{\sigma|n}) \leq \lambda\big(\phi(\sigma|n)\big)\operatorname{diam}(A) \leq \prod_{i=1}^{n} \lambda(w_{\sigma_i})\operatorname{diam}(A)
\tag{2b}
$$

for any $\sigma \in \mathcal{A}$ and any $n \in \mathbb{N}$, where $\lambda(\,.\,) \in [0,1)$. Therefore, the decomposition of μ can be obtained in a recurrent way be means of object instancing used in the measure-theoretic context. Starting with the measure $\mu = \sum_{\sigma_1=1}^{N} \mu_{\sigma_1}$, we replace each sum component nonzero measure $\mu_{\sigma|n}$, for which $\operatorname{diam}(\operatorname{spt}\mu_{\sigma|n}) > \varepsilon$, with m ($m \leq N$) nonzero measures $\mu_{(\sigma|n)\sigma_{n+1}}$, until the diameter of the support of every measure in the sum $\mu = \sum_k \mu_{\sigma_k|n_k}$ does not exceed ε. Such a decomposition into finite sum of measures we will call a ε-decomposition of a measure.

The obvious, yet worth emphasis consequence is that the sum of the total values $\mu_{(\sigma|n)\sigma_{n+1}}(X)$ of the component "sub-measures" $\mu_{(\sigma|n)\sigma_{n+1}}$ is equal to the total value $\mu_{\sigma|n}(X)$ of their "parent" measure. Moreover, if X is a separable metric space, then on the basis (1) we get $\mu_{\sigma|n}(X) = \mu_{\sigma|n}(\operatorname{spt}\mu_{\sigma|n})$, because we deal with Borel measures.

Now, we give examples of types of measures on fractals, which arise from our general definition. We begin with the well-known measures, namely, the IFS and RIFS invariant measures. The standard way of approximating the measures uses the chaos game[1]. However, it is usually hard to estimate the quality of such an approximation in terms of the sufficient number of steps to reach a required approximation accuracy. As a result, we never know whether the reason for an unvisited area of space is the zero measure of the area or too small number of the iterations performed. Another difficulty appears in the case of 3D measures supported, and is due to the inability of incorporating a shading model to the chaos game.

As the second example of fractal measures we present a construction which employs the Perlin noise function[5].

3.1. *IFS and Recurrent IFS invariant measures*

A *recurrent iterated function system*[8] (RIFS) is a generalization of the IFS with probabilities in which instead of a vector of probabilities, a probability transition matrix appears. Formally, a RIFS is a triple $(K; \{w_i\}_{i=1}^N, \mathbf{P})$ so that $(K; \{w_i\}_{i=1}^N)$ is an IFS on a compact metric space K, and $\mathbf{P} = [p_{ik}]$ is a $N \times N$ row-stochastic and irreducible matrix.

The *Markov operator* for the RIFS acts on the space $\widetilde{\mathcal{P}}(K^N)$ of normalized Borel measures on the Cartesian product of N copies of K, and is defined by

$$\widetilde{M}(\widetilde{\nu}) := \Big[\sum_{i=1}^N p_{i1} \nu_i \circ w_1^{-1}, \dots, \sum_{i=1}^N p_{iN} \nu_i \circ w_N^{-1} \Big], \tag{3}$$

where $\widetilde{\nu} = [\nu_1, \dots, \nu_N]$ and ν_i are Borel measures such that $\sum_{i=1}^N \nu_i(K) = 1$. One can show that $\widetilde{M}$ possesses a unique fixed point $\widetilde{\mu} = \widetilde{M}(\widetilde{\mu})$ on a subspace $\{[\nu_1, \dots, \nu_N] : \nu_i(K) = \pi_i, \ i = 1, \dots, N\}$ of $\widetilde{\mathcal{P}}(K^N)$, where $\pi \equiv [\pi_1, \dots, \pi_N] \in \mathbb{R}^N$ is the unique, normalized left-eigenvector of $\mathbf{P}$ [8]. Moreover, the support of $\widetilde{\mu}$ is the vector $[A^{(1)}, \dots, A^{(N)}]$ of compact sets $A^{(i)} \subset K$, such that $A^{(i)} = \bigcup_{k:p_{ki}>0} w_i(A^{(k)})$. If, for each $k \in \{1, \dots, N\}$, $p_{ik} = p_k$ for every $i \in \{1, \dots, N\}$, then the RIFS reduces to the IFS written usually as $(K; \{w_i\}_{i=1}^N; \{p_i\}_{i=1}^N)$. In such a case $\pi = [p_i]_{i=1}^N$.

The *invariant measure* μ generated by the RIFS is a normalized Borel measure on K specified by $\widetilde{\mu}$ as

$$\mu(E) = \widetilde{\mu}[E, \dots, E] = \sum_{i=1}^N \mu_i(E), \quad E \subset K. \tag{4}$$

On the basis of above, μ is supported by the unique compact set

$$A_R = \bigcup_{i=1}^N A^{(i)}, \quad A^{(i)} = \bigcup_{k:p_{ki}>0} w_i(A^{(k)}), \tag{5}$$

that is the attractor of the RIFS[8].

Now we show that μ can be classified to the class of fractal measures arising from our general construction. By definition (4), μ equals the sum of N Borel measures μ_{σ_1}, $\sigma_1 = 1, \dots, N$. Moreover, spt $\mu_{\sigma_1} = A^{(\sigma_1)}$ and $\mu_{\sigma_1}(K) = \pi_{\sigma_1}$, because spt $\widetilde{\mu} = [A^{(1)}, \dots, A^{(N)}]$ and $\widetilde{\mu}$ belongs to the subspace of $\widetilde{\mathcal{P}}(K^N)$ specified by the unique eigenvector π of the matrix $\mathbf{P}$.

Now, each component measure μ_{σ_1} is equal to the sum of m ($m \leq N$) nonzero measures $\mu_{\sigma_1 \sigma_2} \equiv p_{\sigma_2 \sigma_1} \mu_{\sigma_2} \circ w_{\sigma_1}^{-1}$ for which $p_{\sigma_2 \sigma_1} > 0$, $\sigma_2 \in \{1, \dots, N\}$, because $\widetilde{\mu}$ is the fixed point of the operator $\widetilde{M}$. Furthermore, $\mu_{\sigma_2} \circ w_{\sigma_1}^{-1}(w_{\sigma_1}(A^{(\sigma_2)})) \geq \mu_{\sigma_2}(A^{(\sigma_2)})$, because $w_{\sigma_1}^{-1}(w_{\sigma_1}(A^{(\sigma_2)})) \supset A^{(\sigma_2)}$. It follows that spt $\mu_{\sigma_1 \sigma_2} \subset w_{\sigma_1}(A^{(\sigma_2)})$ and $\mu_{\sigma_1 \sigma_2}(K) = p_{\sigma_2 \sigma_1} \pi_{\sigma_2}$, since spt $\mu_{\sigma_2} = A^{(\sigma_2)}$.

In turn, each nonzero measure $\mu_{\sigma_1\sigma_2}$ can be decomposed into the sum of the measures $\mu_{\sigma_1\sigma_2\sigma_3} \equiv p_{\sigma_2\sigma_1}p_{\sigma_3\sigma_2}\mu_{\sigma_3} \circ w_{\sigma_2}^{-1} \circ w_{\sigma_1}^{-1}$, where $p_{\sigma_3\sigma_2} > 0$ and $\sigma_3 \in \{1, \ldots, N\}$, so that $\operatorname{spt}\mu_{\sigma_1\sigma_2\sigma_3} \subset w_{\sigma_1} \circ w_{\sigma_2}(A^{(\sigma_3)})$ and $\mu_{\sigma_1\sigma_2\sigma_3}(K) = p_{\sigma_2\sigma_1}p_{\sigma_3\sigma_2}\pi_{\sigma_3}$.

The above observations lead to the conclusion that each component measure

$$\mu_{\sigma|n} \equiv \left(\prod_{i=1}^{n-1} p_{\sigma_{i+1}\sigma_i}\right)\mu_{\sigma_n} \circ w_{\sigma_{n-1}}^{-1} \circ \ldots \circ w_{\sigma_1}^{-1}, \tag{6}$$

with $\operatorname{spt}\mu_{\sigma|n} \subset \phi(\sigma|(n-1))(A^{(\sigma_n)})$ and $\mu_{\sigma|n}(K) = \left(\prod_{i=1}^{n-1} p_{\sigma_{i+1}\sigma_i}\right)\pi_{\sigma_n}$, of the decomposition of μ can be further decomposed into the sum of m $(m \leq N)$ measures

$$\mu_{(\sigma|n)\sigma_{n+1}} \equiv \left(\prod_{i=1}^{n} p_{\sigma_{i+1}\sigma_i}\right)\mu_{\sigma_{n+1}} \circ w_{\sigma_n}^{-1} \circ w_{\sigma_{n-1}}^{-1} \circ \ldots \circ w_{\sigma_1}^{-1}$$

with $\operatorname{spt}\mu_{(\sigma|n)\sigma_{n+1}} \subset \phi((\sigma|(n-1))\sigma_n)(A^{(\sigma_{n+1})})$ and $\mu_{(\sigma|n)\sigma_{n+1}}(K) = \left(\prod_{i=1}^{n} p_{\sigma_{i+1}\sigma_i}\right)\pi_{\sigma_{n+1}}$, where $p_{\sigma_{n+1}\sigma_n} > 0$ and $\sigma_{n+1} \in \{1, \ldots, N\}$.

The last thing is show that the first condition holds, that is $\operatorname{spt}\mu_{\sigma|n} \subset A_{\sigma|n}$, where $A_{\sigma|n} \equiv \phi(\sigma|n)(A)$ is a subset of the attractor A of the ordinary IFS $(K; \{w_i\}_{i=1}^N)$. By (6) all we have to do is show that $\phi(\sigma|(n-1))(A^{(\sigma_n)}) \subset A_{\sigma|n}$.

Let a be any point of the attractor A_R of a RIFS $(K; \{w_i\}_{i=1}^N, \mathbf{P})$. Then one can show that there is a word $\sigma \in \mathcal{A}$ such that $\lim_{n\to\infty} \phi(\sigma|n)(x) = a$ for any $x \in K$. But $\lim_{n\to\infty} \phi(\sigma|n)(x)$ is also a point of the attractor A of the IFS $(K; \{w_i\}_{i=1}^N)$ [1]. Hence, A_R is a subset of A. In turn, on the basis of (5) we obtain $A^{(i)} = w_i\left(\bigcup_{k:p_{ki}>0} A^{(k)}\right)$, and $\bigcup_{k:p_{ki}>0} A^{(k)} \subset \bigcup_{k=1}^N A^{(k)}$; hence $A^{(i)} \subset w_i(A_R)$. Since $A_R \subset A$, we get $A^{(i)} \subset w_i(A)$, and as a result $\phi(\sigma|(n-1))(A^{(\sigma_n)}) \subset A_{\sigma|n}$. Thus $\operatorname{spt}\mu_{\sigma|n} \subset A_{\sigma|n}$ for any $\sigma \in \mathcal{A}$ and $n \in \mathbb{N}$.

3.2. Perlin-noise measures

Let $(\mathbb{R}^3; \{w_i\}_{i=1}^N)$ be an IFS, and let $\mathcal{N}: \mathbb{R}^3 \to (0, 1)$ be the 3D Perlin noise function [5]. For every $\sigma \in \mathcal{A}$ and $n \in \mathbb{N}^+$, denote by $\mathbf{x}_{\sigma|n}$ the fixed point of the IFS map composition $\phi(\sigma|n)$. Finally, given numbers $\alpha \in \mathbb{R}^+$ and $m \in \{1, \ldots, N\}$, we define a function Γ on all finite (and nonempty) words over $\{1, \ldots, N\}$ by

$$\Gamma((\sigma|n)\sigma_{n+1}) := \begin{cases} \mathcal{N}(\alpha \cdot \mathbf{x}_{(\sigma|n)\sigma_{n+1}}) & \text{if } \mathcal{N}(\alpha \cdot \mathbf{x}_{(\sigma|n)\sigma_{n+1}}) \in G_{\sigma|n} \\ 0 & \text{otherwise} \end{cases},$$

where $\sigma \in \mathcal{A}$, $n \in \mathbb{N}$, $\sigma_{n+1} = 1, \ldots, N$.

Here, $G_{\sigma|n}$ is the set of m greatest values taken by $\mathcal{N}(.)$ for N points $\alpha \cdot \mathbf{x}_{(\sigma|n)\sigma_{n+1}}$, $\sigma_{n+1} = 1, \ldots, N$. (Thus, $0 < |G_{\sigma|n}| \leq m$, so $G_{\sigma|n}$ may be even a singleton—this is the case of $\mathcal{N}(.)$ taking the same value for all N points $\alpha \cdot \mathbf{x}_{(\sigma|n)\sigma_{n+1}}$) Now, using the function Γ, we define μ as a measure that obeys the two conditions of our general construction together with the following additional constraint concerning the total

values of the component measures

$$\mu_{\sigma|n}(\mathbb{R}^3) = \mu_{(\sigma|n)\sigma_{n+1}}(\mathbb{R}^3) \frac{\sum_{\sigma_{n+1}=1}^{N} \Gamma\big((\sigma|n)\sigma_{n+1}\big)}{\Gamma\big((\sigma|n)\sigma_{n+1}\big)}. \tag{7}$$

Clearly, the constraint does not violate the second condition and it is not hard to show that the definition specifies μ in the unique way.

In this definition the numbers α and m play role of free parameters. The former controls the "smoothness" of a measure according to the principle the bigger α the more perturbed measure. The value of the latter reflects in the "lacunarity", i.e. the smaller m the more porous measure.

4. Measure instancing technique

Let $\{X; \{w_i\}_{i=1}^N\}$ be an IFS with the attractor A. Let μ be a Borel measure resulting from the general construction given in Sec. 3. Then, on the basis of the first constraint of the construction and (2), the supports (and their diameters) of the component measures $\mu_{\sigma|n}$ of the decomposition of μ can be approximated with the object instancing for IFS attractors. Starting with a volume enclosing the IFS attractor, a collection of volumes is created using recurrently the principle: If the diameter of a volume is greater than given $\varepsilon > 0$, then the volume is replaced with N "smaller" ones until the diameter of every volume in the collection does not exceed ε. Moreover, the union of volumes in the collection has to include the attractor on every stage of the process of the replacement. Assuming the volume $V_{\sigma|n}$ to include the attractor subset $A_{\sigma|n}$, the new volumes $V_{(\sigma|n)\sigma_{n+1}}$, which are substituted for $V_{\sigma|n}$, are determined on the basis of the IFS map compositions $\phi\big((\sigma|n)\sigma_{n+1}\big)$, so that they enclose the attractor subsets $A_{(\sigma|n)\sigma_{n+1}}$, $\sigma_{n+1} = 1, \ldots, N$.

To approximate the supports of the component measures we use the efficient approach to the object instancing for affine IFS attractors which was proposed in Ref. 7. The method can be regarded as a variant of the adaptive-cut algorithm[9] with modifications to avoid the computation of eigenvalues. Basically, the method utilizes balls centered at the images of the center of an initial ball under the IFS map compositions in a way similar to Ref. 9. However, the radii of the bounding balls are determined as $r = r_0 \sqrt{\sum_i \sum_k m_{ik}^2}$, where $[m_{ik}]$ is the matrix of the linear part of the corresponding map composition and r_0 is the radius of the initial ball that bounds the attractor.

Now we will show how the above observations can be used in the measure-theoretic context so as to approximate the measure of a Borel subset at any accuracy required. Let E be a subset of a separable metric space X. Let μ be a Borel measure on X, which obeys the two conditions given in Sec. 3. Let $\mathcal{B} = \{B_{\sigma_i|n_i}\}$ be a family of (closed) balls obtained by means of the described variant of object instancing, so that $\mathrm{diam}(B_{\sigma_i|n_i}) \leq \varepsilon$, where $\varepsilon > 0$ is a given approximation error allowed. Denote by $\mathcal{B}_E$ the collection of balls $B_{\omega_k|n_k}$ from $\mathcal{B}$ so that for every $q \in \{0, 1, \ldots, n_k\}$ each ball $B_{\omega_k|q}$ intersects E, where $B_{\omega_k|0}$ is a ball bounding the

attractor. Finally, let ψ_ε be the sum of all the component measures $\mu_{\omega_k | n_k}$ whose supports are approximated by the balls from $\mathcal{B}_E$. Clearly, ψ_ε as a finite sum of finite Borel measures on X is a finite Borel measure on X.

Proposition. $\psi_\varepsilon(X)$ *does not increase as* $\varepsilon > 0$ *decreases, and* $\lim_{\varepsilon \to 0} \psi_\varepsilon(X) = \mu(E)$.

Hence, if μ is supported by a subset of $\mathbb{R}^2$, then the measure can be visualized at pixel-size accuracy by computing the μ-measures of pixels and interpreting the obtained values as brightness or color[9].

5. Volume rendering 3D measures

The case of 3D measures requires a more sophisticated approach to rendering. It seems that the method, which is based on volume ray-tracing of gaseous models, is a good choice for this purpose.

5.1. *Brightness equation for measures*

The basis for our illumination model for measures is the volume rendering integral[10] for ray-tracing volume densities. Given a density field $\rho \colon D \to [0, \infty]$, $D \subset \mathbb{R}^3$, the amount of light $I(\mathbf{o}, \vec{\omega})$ coming to the eye $\mathbf{o}$ from direction $\vec{\omega}$, is computed as

$$I(\mathbf{o}, \vec{\omega}) = \int_{t_{\min}}^{t_{\max}} e^{\left(-\tau \int_{t_{\min}}^{t} \rho(\mathbf{r}(u)) du\right)} C(\mathbf{r}(t)) \, \rho(\mathbf{r}(t)) \, dt. \tag{8}$$

Here, $\mathbf{r}(t) = \mathbf{o} + t\vec{\omega}$, $t_{\min} = \max(0, d_{\min})$ and $t_{\max} = \min(d_{\text{global}}, d_{\max})$, so that $\mathbf{r}(d_{\min})$ and $\mathbf{r}(d_{\max})$ respectively are the nearest and the farthest point of the intersection of the ray $\mathbf{r}$ with D, and $\mathbf{r}(d_{\text{global}})$ is the nearest intersection of $\mathbf{r}$ with the rest of the scene. The exp represents the attenuation of light caused by the field, and $C(\mathbf{r}(t))$ is the contribution of light sources to the brightness of a point.

Unfortunately, there are two problems with the use of (8) as a shading model for measures. The first one is due to that the equation operates in terms of density. The natural definition of the density ρ of a measure μ on $\mathbb{R}^n$ specifies ρ to be the Radon-Nikodym derivative of μ with respect to n-dimensional Lebesgue measure, which is usually written $d\mu = \rho dx$, $x \in X \subset \mathbb{R}^n$. However, using the Radon-Nikodym theorem (Refs. 4, 11), one can show that measures on fractals have usually no densities well-defined, at least in the above Lebesgue sense. On the other hand, if μ has a density ρ with respect to a measure φ (in our case the Lebesgue measure), then for every nonnegative measurable function g on X the equality $\int g \, d\mu = \int g\rho \, d\varphi$ holds[11]. This suggests the simple remedy: replace the density factors $d\rho(\mathbf{r}(t))$ and $d\rho(\mathbf{r}(u))$ in (8) with the measure ones $\mu(d\mathbf{r}(t))$ and $\mu(d\mathbf{r}(u))$ respectively to obtain a generalized version of the volume rendering integral.

The second problem is that the integrations are over one-dimensional intervals, whereas the density (and the measure) is 3D. Hence, if it was possible to solve
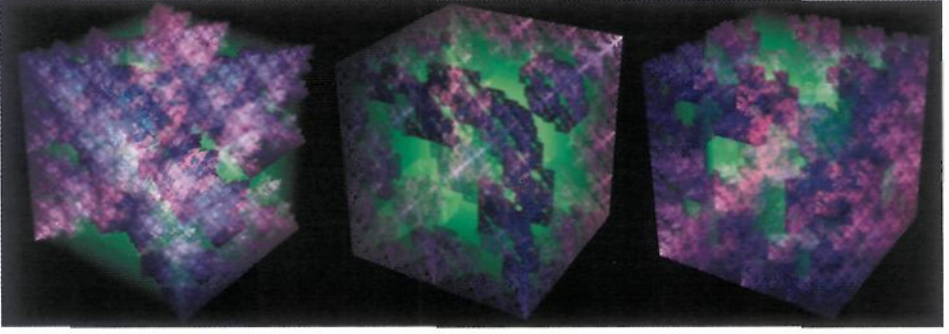
Figure 1.   Results of volume cone tracing RIFS invariant measures supported by subsets of a cube.

the integral in an accurate way, its value would almost always equal 0, because 3-dimensional measure of an interval is usually zero. Consequently, by (8), a density field (and a measure) should not be visible for the observer, at least formally. Therefore, we integrate over 3-dimensional beams of a nonzero volume rather than usual linear rays.

Summarizing, we get the following volume rendering integral for measures. Let μ be a finite measure supported by a compact subset $D \subset \mathbb{R}^3$. Denote by Ω the projection of a given pixel onto the unit sphere that is centered at the eye o serving as the center of projection. The amount of light reaching the eye through the pixel is described by the Lebesgue integral

$$I(\mathbf{o}, \Omega) = \int_{\Omega} \int_{t_{\min}(\vec{\omega})}^{t_{\max}(\vec{\omega})} e^{\left(-\tau \int_{t_{\min}(\vec{\omega})}^{t} \mu(d\mathbf{r}(u,\vec{\omega}))\right)} C(\mathbf{r}(t,\vec{\omega})) \, \mu(d\mathbf{r}(t,\vec{\omega})). \tag{9}$$

Here, $\mathbf{r}(t,\vec{\omega}) = \mathbf{o} + t\vec{\omega}$, and $d\mathbf{r}(t,\vec{\omega})$ is a μ-measurable element of the solid angle $d\omega$ and extent dt with respect to the direction $\vec{\omega}$ [a]. The $C(\mathbf{r}(t,\vec{\omega}))$ factor supplies the light reaching the elementary volume $d\mathbf{r}(t,\vec{\omega})$ as illuminated by light sources.

5.2. Tracing measure with cones

The visualization of a 3D measure relies on evaluating (9) for each pixel of the screen. Assuming rectangular pixels, the area of integration in (9) is a pyramid specified by a pixel and the apex at the eye. Therefore, the integration can be regarded as calculating a measure of the pyramid, so that the μ-measure of each elementary volume $d\mathbf{r}(t,\vec{\omega})$ is weighted by the value of a function $f(t,\vec{\omega})$ representing the attenuation and illumination components. In turn, taking into account the main

---

[a]Points on the unit sphere can be identified with directions. Consequently, $d\omega$ can be considered as an elementary bundle of directions about a direction $\vec{\omega}$, where the size of the bundle is measured in steradians.
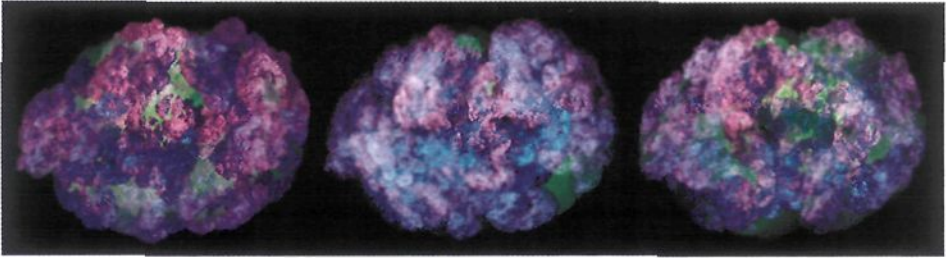
Figure 2.   Examples of cone-tracing of Perlin measures supported by subsets of affine and over-lapping "cloudy" IFS attractor.

external integration over Ω, the integral of the attenuation factor is the μ-measure of an elementary bundle of intervals ranging from the eye to a given element $dr(u, \vec{\omega})$ within the pyramid.

Thus, (9) can be approximated with measure instancing as the "total value" of the measure ψ_ε from Sec. 4. Assuming the set E to be the pyramidal area of integration and identifying the balls from the family $\mathcal{B}_E$ are elementary volume elements, we may compute (9) at accuracy induced by the size of a pixel. The direct application of this idea involves pyramidal rays to look for intersections with balls determined with the object instancing used in the ray-tracing context[6-7]. However, in order to increase the efficiency, rather than pyramids, we utilize cones in the spirit of cone-tracing[12].

Let $\varepsilon > 0$ be an approximation error. In particular, ε may be set that the projection of every ball from $\mathcal{B}$ on the screen does not exceed the width of a pixel. The first step to approximate (9) is to determine family $\mathcal{B}_E = \{B_{\omega_k|n_k}\}$ of $N(\varepsilon)$ balls $B_{\omega_k|n_k}$, $\mathrm{diam}(B_{\omega_k|n_k}) \leq \varepsilon$, intersected by a cone shot through a given pixel. The family specifies the component measures $\mu_{\omega_k|n_k}$ of the sum ψ_ε whose value $\psi_\varepsilon(\mathbb{R}^3) = \sum_{k=1}^{N(\varepsilon)} \mu_{\omega_k|n_k}(\mathbb{R}^3)$ approximates the μ-measure of the cone. In practice, we utilize a list structure, and to compute the balls and values $\mu_{\omega_k|n_k}(\mathbb{R}^3)$ we use measure instancing. We store the results as couples $\langle B_{\omega_k|n_k}, \mu_{\omega_k|n_k}(\mathbb{R}^3) \rangle$ in the list. Next, we sort the list in the increasing order w.r.t the distances of the centers to the eye. Then, proceeding front-to-back, we approximate (9) on the basis of $N(\varepsilon)$ sorted couples $\langle B_i, \mu_i(\mathbb{R}^3) \rangle$ by

$$I(\mathbf{o}, \Omega) \approx \sum_{i=1}^{N(\varepsilon)} e^{(-\tau \sum_{k=1}^{i} \mu_k(\mathbb{R}^3))} C(B_i) \, \mu_i(\mathbb{R}^3). \tag{10}$$

Here, the term $C_\lambda(B_i)$ is the light received by the ith ball from all light sources.

6. Results

Six images of fractal measures are reproduced here to give a flavor of the variety of measures arising from the generic construction proposed in this paper. Besides brightness, we have also computed the average density[4] in balls from $\mathcal{B}$ using the

formula $\widetilde{\rho}(B) = \frac{6}{\pi} \cdot \frac{\mu(B)}{\text{diam}(B)^3}$, so as to reveal the measure's local property. Values $\mu(B)$ have been approximated with measure instancing in parallel with brightness. As it is impossible to represent all the densities along the ray with one color, we have applied the approach used in Ref. 13 and visualize only one value of the discrete density function over the ray—in this work we chose the maximum. To encode the information about density together with the brightness we engage HSV color space and determine the color of a pixel mapping the brightness to the Value component and the maximal density to the Hue component (reversed here so that 'hot' colors indicate high values). Additionally, we utilize the Saturation component to represent the distance between the eye and the maximal density according to the principle the closer to the eye, the more saturated. (We also used symbolically green to mark the attractor subsets of the measure close or equal zero.)

The first three, Fig. 1, seem to be good 3D equivalents of 2D fractal measures whose images serve as a tutorial examples in Ref. 1. They depict examples of RIFS invariant measures supported by a subset of a cube encoded with 8 similarities. Finally, Fig. 2 exemplifies Perlin noise measures carried by a cloudy support specified with 8 affine maps and overlapping IFS. Each of the images was cone-traced at 1200×900 screen resolution and pixel-size accuracy on AMD Athlon 64 Processor 3000+ 2.01 GHz.

Acknowledgments

References

1. M. Barnsley. *Fractals Everywhere, 2nd ed.* Academic Press, 1993.
2. P. Mattila. *Geometry of Sets and Measures in Euclidean Spaces.* Cambridge University Press, 1995.
3. K. Falconer. *Techniques in Fractal Geometry.* Wiley, 1997.
4. G. Edgar. *Integral, Probability, and Fractal Measures.* Springer-Verlag, 1998.
5. K. Perlin. An image synthesizer. *SIGGRAPH '89*, **23**:41–50, 1989.
6. J Hart, and T. DeFanti. Efficient antialiased rendering of 3-D linear fractals. *SIGGRAPH '91*, **25**:91–100, 1991.
7. T. Martyn. Efficient ray tracing affine IFS attractors. *Computers & Graphics*, **25**(4):665–670, 2001.
8. M. Barnsley, J. Elton, and D. Hardin. Recurrent iterated function systems. *Constr Approx*, **5**:3–31, 1989.
9. D. Hepting, P. Prusinkiewicz, and D. Saupe. Rendering methods for iterated function systems. *Fractal '90*, IFIP, 1990.
10. J. Kajiya, and B. von Herzen. Ray tracing volume densities. *SIGGRAPH '84*, **18**:165–174, 1984.
11. W. Rudin. *Real and Complex Analysis, 2nd ed.* McGraw-Hill, 1974.
12. J. Amanatides. Ray tracing with cones. *SIGGRAPH '84*, **18**:129–135, 1984.
13. P. Sabella. A rendering algorithm for visualizing 3D scalar fields. *SIGGRAPH '88*, **22**:51–61, 1988.

COMPLEXITY IN NATURE AND SOCIETY

Complexity Management in the Age of Globalization

KLAUS MAINZER

The t heory o f n onlinear c omplex sy stems h as b ecome a proven p roblem-solving a pproach i n the natural sciences from cosmic and q uantum systems to c ellular organisms and the brain. E ven in modern engineering science self-organizing systems are developed to manage complex networks and processes. It is now recognized that many of our ecological, social, economic, and political problems are also of a global, complex, and nonlinear nature. What are the laws of sociodynamics? Is there a socio-engineering of nonlinear problem solving? What can we learn from nonlinear dynamics for complexity management in social, economic, financial and political systems? Is self-organization an acceptable strategy to handle the challenges of complexity in firms, institutions and other organizations? It is a main thesis of the talk that nature and society are basically governed by nonlinear and complex information dynamics. How computational is sociodynamics? What can we hope for social, economic and political problem solving in the age of globalization?

1. From Linear to Nonlinear Dynamics

A *dynamical system* is characterized by its *elements* and the *time-depending* development of their *states*. In the simple case of a falling stone, a physicist like Galileo only considers the acceleration of a single element. In a planetary system, the states of planets are determined y their position and momentum. The states can also refer to moving molecules in a gas, the excitation of neurons in a neural net, nutrition of populations in an ecological sy stem, o r p roducts i n a market sy stem. T he *dynamics* o f a s ystem, i .e. t he change of system states depending on time, is mathematically described by *differential equations*. For *deterministic processes* (e.g., motions in a planetary system), each future state is uniquely determined by the present state. A *conservative* (Hamiltonian) *system*, e.g. an ideal pendulum, is determined by the reversibility of time direction and conservation of energy. Actually, conservative systems are closed and have no energetic dissipation with t heir e nvironment. T hus, c onservative sy stems i n t he s trict s ense o nly exist as approximations like, e.g., a Thermos bottle. In our everyday world, we mainly observe dissipative systems with a distinct time direction. *Dissipative systems*, e.g., a real pendulum with friction, are irreversible.

In classical physics, the dynamics of a system is considered as a *continuous process*. In a famous quotation, Leibniz assumed that nature does not jump (natura non facit saltus). But, continuity is only a mathematical idealization. Actually, a scientist has

single observations or measurements at discrete-time points which are chosen equidistant or defined by other measurement devices. In discrete processes, there are finite differences between the measured states and no infinitely small differences between the measured states and no infinitely small differences (differentials) which are assumed in a continuous process. Thus, discrete processes are mathematically described by *difference equations*.

Random events (e.g., Brownian motion in a fluid, mutation in evolution, innovations in economy) are represented by additional *fluctuation terms*. *Classical stochastic processes*, e.g. the billions of unknown molecular states in a fluid, are defined by time-depending differential equations with distribution functions of probabilistic states. In *quantum systems* of elementary particles, the dynamics of quantum states is defined by Schrödinger's equation with observables (e.g., position and momentum of a particle) depending on *Heisenberg's principle of uncertainty* which allows only *probabilistic forecasts* of future states.

Historically, during the centuries of classical physics, the universe was considered as a deterministic and conservative system. The astronomer and mathematician P.S. Laplace (1814), e.g., assumed the total computability and predictability of nature if all natural laws and initial states of celestial bodies are well known. The *Laplacean spirit* expressed the belief of philosophers in determinism and computability of the world during the 18th and 19th century.

Laplace was right about *linear* and *conservative dynamical systems*. A simple example is a so-called harmonic oscillator like a mass attached to a spring oscillating regularly without friction. This example is considered in more detail in order to introduce the basic notions of time series, phase space and trajectory. Only with these instruments it is possible to recognize and to understand the structure and development of dynamical systems. In general, *a linear relation* means that the rate of change in a system is proportional to its cause: Small changes cause small effects while large changes cause large effect. In the example of a harmonic oscillator, a small compression of a spring causes a small oscillation of the position of a mass, while a large compression causes a large oscillation according to Hooke's law.

Changes of a dynamical system can be modeled in one dimension by changing values of a time-depending quantity along the time axis (*time series*). For example, the position $x(t)$ of a mass attached to a spring is oscillating in regular cycles along the time axis t. $x(t)$ is the solution of a linear equation, according to Hooke's law. Mathematically, linear equations are completely computable. This is the deeper reason for Laplace's philosophical assumption to be right for linear and conservative systems.

In systems theory, the *complete information* about a dynamical system at a certain time is determined by its *state* at that time. In the example of an harmonic oscillator, the state of the system is defined by the position $x(t)$ and the velocity $v(t)$ of the oscillating mass at time t. Thus, the state of the system is completely determined by a pair of two quantities which can be geometrically represented by a point in a 2-dimensional phase space with a coordinate of position and a coordinate of velocity. The dynamics of a system means the time-depending development of its states. Thus, the dynamics of a system is illustrated by an orbit of points (*trajectory*) in a phase space corresponding to the time-depending development of its states. In the case of an harmonic oscillator, the trajectories are closed ellipses around a point of stability, corresponding to the periodic cycles of time series, oscillating along the time axis. Obviously the regular behavior of a

linear and conservative system corresponds to a regular and stable pattern of orbits. So, the past, presence, and future of the system is completely known.

In general, the state of a system is determined by more than two quantities. Then, a higher dimensional phase space is needed to study the dynamics of a system. From a methodological point of view, time series and phase spaces are important instruments to study systems dynamics. The *state space of a system* contains the *complete information of its past, present and future behavior*. The dynamics of real systems in nature and society is, of course, more complex, depending on more quantities with patterns of behavior which are not as regular as in the simple case of an harmonic oscillator. It is a main insight of modern systems theory that the behavior of a dynamic system can only be recognized if the corresponding state space can be reconstructed.

At the end of the 19th century, H. Poincaré (1892) discovered that celestial mechanics is not a completely computable clockwork, even if it is considered as a deterministic and conservative system. The mutual gravitational interactions of more than two celestial bodies ('Many-bodies-problem') correspond to *nonlinear* and *non-integrable equations* with *instabilities* and *irregularities*. According to the Laplacean view, similar causes effectively determine similar effects. Thus, in the phase space, trajectories that start close to each other also remain close to each other during time evolution. Dynamical systems with *deterministic chaos*, exhibit an exponential dependence on initial conditions for bounded orbits: the separation of trajectories with close initial states increases exponentially.

Thus, tiny deviations of initial data lead to exponentially increasing computational efforts for future data limiting long-term predictions, although the dynamics is in principle uniquely determined. This is known as the 'butterfly effect': initial, small and local causes soon lead to unpredictable, large and global effects (see Fig. 4). According to the famous KAM-Theorem of A.N. Kolmogorov (1954), V.I. Arnold (1963), and J. K. Moser (1967), trajectories in the phase space of classical mechanics are neither completely regular, nor completely irregular, but depend sensitively on the chosen initial conditions.

Dynamical systems can be classified on the basis of the effects of the dynamics on a region of the phase space. A conservative system is defined by the fact that, during time evolution, the volume of a region remains constant, although its shape may be transformed. In a *dissipative system*, dynamics causes a *volume contraction*.

An *attractor* is a region of a phase space into which all trajectories departing from an adjacent region, the so-called *basin of attraction*, tend to converge. There are different kinds of attractors. The simplest class of attractors are the *fixed points*. In this case, all trajectories of adjacent regions converge to a point. An example is a dissipative harmonic oscillator with friction: the oscillating system is gradually slowed down by frictional forces and finally come to a rest in an equilibrium point.

Conservative harmonic oscillators without friction belong to the second class of attractors with *limit cycles*, which can be classified as being periodic or quasi-periodic. A *periodic orbit* is a closed trajectory into which all trajectories departing from an adjacent region converge. For a simple dynamical system with only two degrees of freedom and continuous time, the only possible attractors are fixed points or periodic limit cycles. An example is a Van der Pol oscillator modelling a simple vacuum-tube oscillator circuit.

In continuous systems with a phase space of dimension $n > 2$, more complex attractors are possible. Dynamical systems with *quasi-periodic limit cycles* show a time

evolution which can be decomposed into different periodic parts without a unique periodic regime. The corresponding time series consist of periodic parts of oscillation without a common structure. Nevertheless, closely starting trajectories remain close to each other during time evolution. The third class contains dynamical systems with *chaotic attractors* which are *non-periodic*, with an *exponential dependence on initial conditions* for *bounded orbits*. A famous example is the chaotic attractor of a Lorenz system simulating the chaotic development of weather caused by local events, which cannot be forecast in the long run (*butterfly effect*).

From a methodological point of view, we have started by seeing the kind of mathematical equations of dynamical systems required to derive their patterns of behaviour; the latter have been characterized by time series and attractors in phase spaces, such as fixed points, limit cycles, and chaos. This *top-down approach* is typically theoretical: we use our understanding of real systems to write dynamical equations. In empirical practice, however, we must take the opposite *bottom-up approach* and start with finite sequences of measurements, i.e. finite time series, in order to find appropriate equations of mathematical models with predictions that can be compared with measurements made in the field of application.

Measurements are often contaminated by unwanted noise which must be separated from the signals of specific interest. Further on, in order to forecast the behaviour of a system, the development of its future states must be reconstructed in a corresponding phase space from a finite sequence of measurements. Thus, *time-series analysis* is an immense challenge in different fields of research from, e.g., climatic data in meteorology, ECG-signals in cardiology, and EEG-data in brain research to economic data of economics and finance.

In principle, we distinguish *linear* and *nonlinear stochastic models* of time-series analysis. Being linear, their properties can be fully and rigorously derived from their model equations. In order to generate *aperiodic* time series data, such models require stochastic inputs. Auto regressive (AR) models with Gaussian distributed white noise represent the time discretized motion of the superposition of damped harmonic oscillators driven by noise under some stability conditions. Thus, spectral analysis is the most suitable analysis tool. The sunspot number time series of solar activity with its pronounced 11-year period can be captured by an AR-model. AR-models are suitable for data sets that have a few pronounced peaks in their power spectrum. If the the observed power spectrum instead is broad band, the moving average (MA) model can be a more appropriate tool. A combination of both, the ARMA-models, describes many noise-dominated signals. A different class of sources for aperiodic time-series data arises from chaotic dynamical systems. In this case, aperiodicity comes from intrinsic instabilities without random inputs. *Nonlinearity* destroys the superposition principle, such that a Fourier decomposition of the signal is not useful. If one assumes that a given experimental observable represents the deterministic dynamics in some non-observed and even unknown phase space, the concept of phase space reconstruction by embedding is needed.

In practice, only a time series of a single (one-dimensional) measured variable is often given, although the real system is multidimensional. The aim of forecasting is to predict the future evolution of this variable. According to Takens' theorem (1981), in nonlinear, deterministic and chaotic systems, it is possible to determine the structure of the *multidimensional dynamic system* from the measurement of a *single dynamical*

variable. Takens' method results in the construction of a multidimensional *embedding phase space for measured data* with a certain time delay in which the dynamics of attractors is similar to the orbits in the phase space of the chaotic system.

The disadvantage of Takens' theorem is that it does not detect and prove the existence of a chaotic attractor. It only provides features of an attractor from measured data, if the existence of the attractor is already guaranteed (Grassberger 1983). The *dimension of an attractor* can be determined by a *correlation integral* defining the different frequency with which a region in an attractor is visited by the orbits. Thus, the correlation integral also provide a method to study the *degrees of periodicity and aperiodicity* of orbits and measured time series.

The *Lyapunov Spectrum* shows us the *dependence of dynamics from initial data.* The Lyapunov exponents measure the averaged exponential rates of divergence or convergence of neighbouring orbits in phase space. If the largest Lyapunov exponent is positive, the attractor is *chaotic*, and the initial small difference between two trajectories will diverge exponentially. If the largest exponent is zero and the rest is negative, the attractor is a *periodic limit cycle*. If there is more than one exponent equal to zero, the rest being negative, the behaviour is *quasi-periodic*. If the exponents are all negative, the attractor is a *fixed point*. In general, for *dissipative systems*, the sum of Lyapunov exponents is negative, in spite of the fact that some exponents could be positive.

2. Complexity and Nonlinear Dynamics of Evolution and Brains

Structures in nature can be explained by the *dynamics* and *attractors* of complex systems. They result from collective patterns of interacting elements that cannot be reduced to the features of single elements in a complex system. *Nonlinear interactions* in multi-component ('complex') systems often have synergetic effects which can neither be traced back to single causes nor be forecast in the long run. The mathematical formalism of complex dynamical systems is taken from statistical physics. But, in general, the theory of complex dynamical systems deals with profound and striking analogies which have been discovered in the self-organized behaviour of quite different systems in physics, chemistry, and biology. These multi-component systems consist of many units like elementary particles, atoms, cells or organisms. The elementary units, e.g., their position and momentum vectors, and their local interactions constitute the microscopic level of description, for instance, the interacting molecules of a liquid or gas. The global state of the complex systems results from the collective configurations of the local multi-component states. At the macroscopic level, there are few collective ('global') quantities like, for instance, pressure, density, temperature, and entropy characterizing observable collective patterns or figures of the units.

If the external conditions of a system are changed by varying certain control parameters (e.g., temperature), the system may undergo a change in its macroscopic global states at some threshold value. For instance, water as a complex system of water molecules changes spontaneously from a liquid to a frozen state at the critical value of temperature with zero Celsius. In physics, those transformations of collective states are

called *phase transitions*. Obviously, they describe a change of self-organized behavior between the interacting elements of a complex system.

According to L. D. Landau (1959), the suitable macrovariables characterizing this change of global order are denoted as *'order parameters'*. In statistical mechanics the order transition of complex systems like fluids, gases, etc. is modeled by differential equations of the global state. A paradigmatic example is a ferromagnet consisting of many elementary atomic magnets ('dipoles'). The two possible local states of a dipole are represented by upwards and downwards pointing arrows. If the temperature (*'control parameter'*) is annealed to the thermal equilibrium which is in this case the Curie point, then the average distribution of upwards and downwards pointing dipoles (*'order parameter'*) is spontaneously aligned in one regular direction. This regular pattern corresponds to the macroscopic state of magnetization. Obviously, the emergence of magnetization is a self-organized behaviour of atoms which is modelled by a phase transition of a certain order parameter, the average distribution of upwards and downwards pointing dipoles.

Landau's scheme of phase transitions cannot be generalized to all cases of phase transitions. A main reason for its failure results from an inadequate treatment of *fluctuations*, which are typical for many multi-component systems. Nevertheless, Landau's scheme can be used as a heuristic device to deal with several *non-equilibrium transitions*. In this case, a complex system is driven away from equilibrium by increasing energy (not decreasing energy as in the case of equilibrium transitions like freezing water or magnetizing ferromagnets). The phase transitions of nonlinear dissipative complex systems far from thermal equilibrium can be modelled by several mathematical methods.

As an example, consider a *solid-state laser* which consists of a set of laser-active atoms embedded in a solid-state configuration. The laser end faces act as mirrors serving two purposes: they select light modes in axial direction and with discrete frequencies. If the laser atoms are excited only weakly by external sources ('*control parameters*'), the laser acts as an ordinary lamp. The atoms independently of each other emit wave-tracks with random phases. The atoms, visualized as oscillating dipoles, are oscillating completely at random. If the level of excitement is further increased, the atomic dipoles spontaneously oscillate in phase, although they are excited completely at random. Obviously, the atoms show a *self-organized behaviour of great regularity*. The extraordinary coherence of laser light results from the collective cooperation of the atomic dipoles.

The laser shows *features of phase transitions*. Order parameters describe mode amplitudes of the light field becoming *unstable* at a critical value of pumping. These slowly varying amplitudes now *"slave"*, as H. Haken claimed, the atomic system during a critical transition. The atoms have to "obey" the orders of order parameters. This mathematical scheme has a very comfortable consequence: it is not necessary (and not possible) to compute all microstates of atoms in a complex system, find the few macroscopic order parameters, and you understand the dynamics of a complex system.

Actually, the corresponding equations describe a competition of several order parameters among each other. The atoms will then obey that order parameter that wins the competition. A typical example is a *Bénard experiment* analyzing the emergence of convection rolls in a fluid layer at a critical value of a control parameter (temperature). Examples can also be observed in the layers of the atmosphere. In this case, the order parameters correspond to the two possible rolling directions "left" or "right" of the

convection rolls. During the *phase transition* of increasing temperature it cannot be forecast which of the two possible order parameters will win the competition, because it depends on tiny initial fluctuations on the molecular level. Thus, this phase transition corresponds to a *spontaneous symmetry breaking* of two possible orders. *Fluctuations* are the driving forces of the system's evolution.

In a more qualitative way we may say that old structures become unstable, broken down by changing control parameters, and *new structures* and attractors are achieved. If, for example, the fluid of a stream is driven further and further away from thermal equilibrium, e.g., by increasing fluid velocity (control parameter), then fluid patterns of increasing complexity emerge from *vortices of fixed points, periodic* and *quasi-periodic oscillations* to *chaotic turbulence.*

In a more mathematical way, stochastic nonlinear differential equations (e.g., Fokker-Planck-equations, Master equation) are employed to model the dynamics of complex systems. The dominating order parameters are founded by the adiabatic elimination of fast relaxing variables of these equations. The reason is that the relaxation time of unstable modes (order parameters) is very long compared to the fast relaxing variables of stable ones, which can therefore be neglected. Thus, this concept of self-organization can be illustrated by the quasi-biological slogan: long-living systems dominate short-living systems.

Dynamical systems and their *phase transitions* deliver a successful formalism to model the *emergence of order in nature.* But these methods are not reduced to special laws of physics, although their mathematical principles were first discovered and successfully applied in physics. There is *no physicalism,* but an interdisciplinary methodology to explain the increasing complexity and differentiation of forms by phase transitions. The question is how to select, interpret and quantify the appropriate variables of dynamical models. Let us consider a few examples.

Thermodynamic self-organization is not sufficient to explain the *emergence of life.* As nonlinear mechanism of genetics we use the autocatalytic process of genetic self-replication. The evolution of new species by mutation and selection can be modelled by nonlinear stochastic equations of second-order nonequilibrium phase transitions. Mutations are mathematized as 'fluctuating forces' and selections as 'driving forces'. Fitness degrees are the order parameters dominating the phase transitions to new species. During e volution a sensible network o f e quilibria b etween p opulations o f a nimals a nd plants has developed. The nonlinear Lotka-Volterra equations model the ecological equilibrium between prey and predator populations which can be represented by oscillating time series of population growth or limit cycles around points of stability. Open dissipative systems of ecology may become unstable by local perturbations, e.g., pollution of the atmosphere, leading to global chaos of the atmosphere in the sense of the *butterfly effect.*

In brain research, *brain* is considered a complex dynamical system of firing and non-firing neurons, self-organizing in macroscopic patterns of cell assemblies by neurochemical interactions. Their dynamical attractors are correlated with states of perception, motion, emotion, thoughts, or even consciousness. There is no 'mother neuron' which can feel, think or at least coordinate the appropriate neurons. The famous *binding problem* of pixels and features in a perception is explained by clusters of synchronously firing neurons dominated by learnt attractors of brain dynamics. The brain is also a self-monitoring and self-mapping system of all bodily, cognitive, and emotional

states leading to *self-awareness* and *self-consciousness* which can be interpreted as dominating order parameters. Thus, even human subjectivity, the traditional philosophical problem of 'qualia', can be explained by nonlinear dynamics of complex systems. H uman i ntentions a nd p references c orrespond t o a ttractors o f b rain d ynamics influencing human acting and behaviour.

In general, complex dynamical systems can be characterized by a specific time scaling. I. Prigogine draws a distinction between the reversible '*exterior*' time or its '*age*'. While the 'exterior' time is the usual real time parameter t that is registered by a clock, the '*interior*' time is defined as an operator that takes into account the irreversible changes in the system's states. As a real parameter, the exterior time appears merely as an index in a unique set of trajectories (in classical physics) or in a wave equation (in quantum mechanics). As an operator, the interior time permits statements about the temporal development of a complex collection of trajectories or distribution functions that serve mathematically as the eigenfunctions of the time operator. The connection with external time rests on the eigenvalues of the time operator being real length of time as registered by a normal clock. The distributions represent the different interior 'ages' of a complex system. For example, the different organs of a complex system like the human organism wear o ut a t d ifferent st ates. T he t ime o perator a ssigns a 'mean a ge' t o e ach state of the system, which increases at the same rate as the exterior clock time. In this sense, each dynamical system has its interior time. A well-known example is the subjectively experienced time of consciousness. The subjectively experienced *duration of a time* depends on the frequency and distribution of occurrences which are registered by the brain in the corresponding period of clock time. In periods of business, time seems to speed up. But during periods without action, time runs slowly.

3.   Complexity and Nonlinear Dynamics of Economies and Finance

The self-organization of complex systems can also be observed in social groups. If a group of workers is directed by another worker, the so-called foreman, then we get an organized behaviour to produce some product which is by no means self-organized. Self-organization means that there are no external orders of a foreman, but the workers work together by some kind of mutual understanding, each one doing his job according to a collective concept dominating their behaviour. Another application of social dynamics is the behaviour of car drivers. In *automobile traffic systems*, a phase transition from non-jamming to jamming phases depends on the averaged car density as control parameter. The spontaneous emergence of chaotic patterns of traffic is a famous self-organizing effect of nonlinear interactions which can often not be reduced to single causes. At a critical value, fluctuations with fractal or self-similar features can be observed. The term *self-similarity* states that the time series of measured traffic flow looks the same on different time scales, at least from a qualitative point of view with small statistical deviations. This phenomenon is also called *fractality*. In the theory of complex systems, self-similarity is a (not sufficient) hint on chaotic dynamics. These signals can be used by guide systems of traffic.

In a political community, collective trends or majorities of opinions can be considered as *order parameters* which are produced by mutual discussions and

interaction of the people in a more or less 'heated' situation. They can even be initiated by some few people in a critical and unstable ('revolutionary') situation of the whole community. There may be a competition of order concepts during heavy fluctuations. The essential point is that the winning concept of order will dominate the collective behaviour of the people. Thus, there is a kind of feedback: the collective order of a complex system is generated by the interactions of its elements (*'self organization'*). After thermodynamic, genetic, and neural self-organization we also distinguish *social* and *economic self-organization*. On the other hand, the behaviour of the elements is dominated by the collective order. People have their individual will to influence collective trends of society. Nevertheless, they are also driven by attractors of collective behaviour.

But, in contrast to physical, chemical, and biological systems, no equations of motion on the microlevel are available for social systems. People are not atoms or molecules, but human beings with intentions, motivations, and emotions. In principle, their individual behaviour and decision-making could be explained by their *brain dynamics*. Cognitive and emotional dynamics are determined by order parameters characterizing individual thoughts, decisions, and motivations. But this is only a theoretical option, because the corresponding neural equations are not yet known. Furthermore, they would be too complex to solve them and predict the future behaviour of people.

Therefore an alternative approach is suggested which gets along without microscopic equations, but nevertheless takes into account the decisions and actions of individuals with probabilistic methods in order to derive the *macrodynamics of social systems*. The modelling design consists of three steps: In the first step appropriate variables of social systems must be introduced to describe the states and attitudes of individuals. The second step defines the change of behaviour by probabilistic phase transitions of individual states. The third step derives equations for the global dynamics of the system by stochastic methods.

In a society we can distinguish several sectors and sub-sectors that are denoted by variables. There are variables for material states, extensive and intensive personal states. The socioconfiguration of a social system is characterized by these material and personal macrovariables. They are measured by usual methods of demoscopy, sociology, or economics. Like in thermodynamics, there are intensive economic variables that are independent of the size of a system. Examples are prices, productivity, and the density of commodities. Extensive variables are proportional to the size of a system and concern, for example, the extent of production and investment, or the size and number of buildings. Collective material variables are measurable. Their values are influenced by the individual activities of agents, which are often not directly measurable. The social and political climate of a firm is connected to socio-psychological processes, which are influenced by the attitudes, opinions, or actions of individuals and their subgroups. Thus, in order to introduce the socioconfiguration of collective personal variables, we must consider the states of individuals, expressed by their attitudes, opinions, or actions. Furthermore, there are subgroups with constant characteristics (e.g., sections or departments of a firm or an institution), so that each individual is a member of one subgroup. The number of members of a certain state is a measurable macrovariable. The socioconfiguration of, e.g., a company is a set of macrovariables describing the distribution of attitudes, opinions, and actions among its subgroups at a particular time.

The total macroconfiguration is given by the multiple of material configuration and socioconfiguration.

If all macrovariables of a macroconfiguration remain constant over time, the social system is in a *stationary macroscopic equilibrium*, which can be compared to thermodynamic equilibrium. If there are dynamics, we must consider the *transition rate* between macroconfigurations by either increasing or decreasing macrovariables. In the case of material configuration, an elementary change consists of the increase or decrease of one macrovariable (e.g., the price of commodities) by one appropriately chosen unit. The elementary change in the socioconfiguration takes place if one individual of a subgroup changes its state, leading to an increase or decrease in the number of a subgroup by one unit. For example, the variable of employment is diminished or enlarged by one person or a political voter changes his preferred party in a certain period of time. These transitions of individual states should be described by a probabilistic process, because the individual freedom of decision and action should not be restricted. People are not molecules with deterministic microdynamics. Nevertheless, the phase transitions must take into account running trends and motivations in order to get a realistic estimate of the probabilities. These aspects provide the link between the microlevel of individual decisions and the macrolevel of collective behaviour. There are well-known statistical procedures of data-mining to measure probabilistic trends, motivations, and attitudes of people belonging to certain sections and subsections of a society.

The *probabilistic phase transitions* can be used for setting up the macroevolution equation of a social system. The probabilistic macrobehavior of a society is described by a probability distribution function over its possible socioconfigurations at a certain time. The distribution function $P(m, n; t)$ can be interpreted as the probability of finding a certain macroconfiguration of material configuration m and socioconfiguration n at time t. The evolution of the social system is the time-depending change of its probabilistic macrobehavior, i.e. the time derivative of the probability function $dP(m, n; t)/dt$. Thus, we get a stochastic nonlinear differential equation which is well-known in thermodynamics as *master equation*. In natural science (e.g. thermodynamics), the evolution of a whole ensemble of millions of equally prepared but probabilistically evolving systems (e.g. gas, fluids) can often be measured. Therefore, the master equation with its probabilistic distribution function of macrostates is appropriate in natural sciences, although only numerical solutions are in general available. This method of modelling macrodynamics is called *ensemble approach*.

But social sciences (e.g. economics, politics, sociology) only dispose of one or at best of a few comparable systems. Therefore, the probability distribution of the master equation contains too much information in comparison to the available empirical data. Sociodynamics focuses on the *stochastic evolution* of a single system in which it traverses probabilistically a sequence of system states. In the corresponding state space of socioconfigurations, we get stochastic trajectories describing the probabilistic dynamics of social systems. The stochastic evolution of a single social system is determined by autonomous nonlinear differential equations for the stochastic trajectories of material, extensive and intensive personal variables. In this sense, the emergence of social structures and patterns of behaviour can again be modelled by solutions of nonlinear equations.

An example of social phase transitions and symmetry breaking is provided by world-wide *migration processes*. The behaviour and the decisions of people to stay or to

leave a region are illustrated by spatial distributions of populations and their change. The models may concern regional migration within a country, motivated by different economic and urban developments, or even the dramatic worldwide migration between poor and rich countries in the age of globalization. The migration interaction of two human populations may cause several macro-phenomena, such as the emergence of a stable mixture, the emergence of two separated, but stable ghettos, or the emergence of a restless migration process. In numerical simulations and phase portraits of the migration dynamics, the macro-phenomena can be identified with corresponding attractors. The empirical administrative data can be used to test the models.

Another model of two *competing firms* provide a bifurcation into winner and loser at a critical value of competition. In nonlinear dynamics, order parameters of the macrodynamics of a system are introduced by linear-stability analysis. The idea of the model is that quality is the order parameter of competition dominating all other economic aspects. Thus, parameters of, e.g., prices, supply, or purchase activities of customers can be neglected, and we get an macroevolution equation of quality. The evolution of the quality variables q_i for firm $i = 1$ and firm $i = 2$ can be investigated numerically. Their (stationary) solutions $q_i(\phi)$ are depicted as a function of the competition parameter ϕ. It turns out that both firms have the same stationary quality $q(\phi)$ of their products and also the same stationary market share, as long as the competition value ϕ is smaller than a critical value ϕ_c. At ϕ_c, a bifurcation occurs and for $\phi > \phi_c$ there exist two stable quality values $q_+(\phi)$ and $q_-(\phi)$. The winning firm, say $i = 1$, will have reached the quality $q_+(\phi)$, whereas the losing firm, $i = 2$, arrives at quality $q_-(\phi)$ with corresponding market shares.

Economic dynamics is concerned with fluctuations in the economy. Most economic variables exhibit perpetual fluctuations over time. These fluctuations are characterized by growth of production and employment as well as large oscillations in relative changes or growth rates The fluctuations vary from more or less regular business cycles in macroeconomic variables to very irregular fluctuations, e.g., in stock prices and exchange rates, in financial markets. Again, we distinguish linear and nonlinear explanations of the fluctuations. According to the traditional new classical point of view, the main source of fluctuations is to be found in exogenous, random shocks to an inherently stable, often linear economic system. Without any external shocks and disturbances, the economy would be stable and converge to the unique steady-state path. Actually, simple linear systems buffeted with noise could mimic time series similar to those in real business cycle data. According to the opposing point of view, economic fluctuations are not caused by random impulses, but should be explained by nonlinear economic laws. It is an old Keynesian topic that fluctuations can also be driven by volatile, self-fulfilling expectations of market psychology.

In the 1940s and 1950s N. Kaldor, J. Hicks, and R. Goodwin developed nonlinear dynamic models with locally unstable steady states and stable limit cycles as an explanation for business cycles. But these early nonlinear models were criticized because of forecasting errors. Further on, from a dogmatic point of view, they suffered from the fact that their laws of behaviour were not derived from the traditional belief in rationality in the sense of utility and profit maximizing principles. Inspired by chaos theory, economists have looked for low deterministic chaos in economic and financial data like in natural sciences. But methods like the correlation dimension test need very long time series and depend very sensitively on noise. Therefore, the evidence of low-dimensional

deterministic chaos in economic and financial data is still low, but nonlinearity in general is an increasing challenge.

Contrary to dynamical systems of the natural sciences, an economy is an expectation feedback system. Contrary to, e.g., molecules in a gas, decisions of economic agents are based on their expectations and beliefs about the future state of the economy. Through these decisions, expectations feed back into the economy and affect several economic variables. In the stock market, e.g., this process may lead to a self-fulfilling speculative bubble. Nonlinear dynamics, the possibility of wild fluctuations, turbulence, chaos and its limitations of predictability shed new light on the expectations hypothesis. In a linear stable economy, agents can learn to have rational expectations in the long run. But in a complex nonlinear world, with prices and quantities moving irregularly on chaotic attractors and sensitivity to initial conditions, a boundedly rational world view seems to be more appropriate. Obviously, there are rational and nonrational agents. Therefore, heterogeneous agents are used in, e.g., financial market modelling. In this case, an economy is considered a complex evolving system of many different, boundedly rational, interacting traders with strategies, expectations realizations co-evolving over time. Fundamentalists base their investment decisions upon market fundamentals such as dividends, growth indicators, etc., but chartists only pay attention to patterns and trends in past prices. An evolutionary competition between them may lead to irregular switching between the different strategies with the result of price fluctuations. In evolutionary economies in the Schumpeterian sense, adaptive learning and innovations are also possible. Finally, nobody knows the economic solution: nonlinear dynamics decides who is right like in natural evolution.

Nevertheless, the question arises: what can we learn from time series analysis of, e.g., financial data which deliver the measure of our political and economic decisions. Stability and welfare of our societies depend sensitively on the dynamics of international financial markets. We already mentioned that, in general, we do not know the microscopic motions of economic data and agents. Therefore, in 1900, the French mathematician L. Bachelier considered the fluctuations of stock prices as statistical random walk (Brownian motion) before physicists like A. Einstein (1905) discovered it in the microscopic motion of small particles in fluids. Brownian motion does not only imply statistical stationarity of price increments and scaling of prices (i.e. invariance under displacement and change of scale), but also independence of price increments (knowing the past brings no knowledge about the future), continuity of price variation (a sample of Brownian motion is a continuous curve), rough evenness of price changes (normal Gaussian distribution or "white noise"), absence of clustering (no emergence of local patterns and structure) and absence of cyclic behaviour. Further on, Gaussian distribution leads to the assumption of an efficient market and successful arbitraging with prices following a martingale: whether the past is known in full, in part, or not at all, price changes over all future time spans have zero as expectation.

Modern theory and practice of finance is still more or less based upon these fundamental beliefs. What makes an investment in the stock market risky is that there is a spread of possible outcomes. The usual measure of this spread is believed to be the standard deviation or variance in a bell-shaped (Gaussian) normal distribution. On this basis, H. Markowitz (1952) suggested his well-known construction of portfolios in order to diversify away unique risk. A stock's contribution to the risk of a fully diversified portfolio depends on its sensitivity to market changes which is measured by a parameter

'beta'. Beta delivers the benchmarks for the expected risk premium which, since the mid-1960s, was calculated by the capital asset pricing model (CAPM). Bachelier did not only suggest the random walk of price changes, but also considered the effects of investing in options. A breakthrough in 1973 was the famous Black-Scholes formula calculating the present call option when there is a continuum of possible future stock prices on the basis of a normal (Gaussian) distribution. Every day dealers on the options exchanges still use this formula to make their trades.

Brownian motion is mathematically more manageable than any alternative. But, unfortunately, it is an extremely poor approximation to financial reality. Since the end of the 1980s, we can observe financial crashes and turbulences deviating significantly from normal distribution. Investment portfolios collapsed and hedging with options à la Black-Scholes failed. From the viewpoint of dynamical systems, patterns of time series analysis illustrate the failures of traditional financial theory. While a record of Brownian motion changes looks like a kind of 'grass' with normal length, a record of actual price changes looks like an irregular alternation of quiet periods and bursts of volatility that stand out from the normal length of the grass. This feature demonstrates the apparent non-stationarity of the underlying rules. Further on, discontinuities appear as sharp peaks from the normally distributed Gaussian 'grass'. These peaks are not isolated but bunched together. Cyclic (but not periodic) behaviour can be observed. Instability of the sample variance is expressed by a long-tailed distribution price changes. Last but not least, there is a long-term dependence of data.

Financial markets display some common properties with fluid turbulence. As for fluid turbulent fluctuations, financial fluctuations have intermittency at all scales. In fluid turbulence, a cascade of energy flux is known to occur from the large scale of injection to the small scales of dissipation. Since the 1980s, these cascades lead to multifractal fields. In finance, the picture of a cascade of information flowing from large-scale investors to small scale ones has been proposed. The fluctuations of various financial time series possess multifractal statistics, abandoning the traditional Brownian motion picture. The financial system is no longer a linear, continuous and rational machinery, but characterized by nonlinearities and multifractals. B.B. Mandelbrot reflected two stories in the Bible, those of the flood (Noah effect) and the Seven Fat and seven Lean Cows (Joseph effect), in order to illustrate discontinuous jumps and long dependence. He suggested several parameters (e.g., Hurst coefficient) and criteria (rescaled range analysis R/S) to measure these effects statistically.

In the nonlinear and fractal approach of the financial system randomness can no longer be restricted to the 'normal' Gaussian distribution of price changes. We must consider degrees of randomness. Gaussian distribution corresponds to a pattern of time series with a 'normal' length of 'grass' without extreme peaks. Therefore, it is called mild randomness which can be compared to a solid state of matter aggregation with low energy, stabile structure and defined volume. Wild randomness resembles the gas phase of matter with high energy, less structure and no defined volume. Slow randomness means the fluid state between the gas and solid state. From the viewpoint of time series, mild randomness corresponds to short- and long-run evenness. Slow randomness corresponds to short-run concentration and long-run evenness. Wild randomness corresponds to short- and long-run concentration.

The financial system is not only nonlinear and fractal, but has a typical trade time. It is the interior time of the financial system which must be distinguished from the

'exterior' time of clocks. Professional traders talk about 'slow' and 'rapid' markets depending on their calculation of volatility: in periods of high volatility and business, time seems to run. But during low volatility, time seems to be slow without occurrences and action. It is remarkable that the psychological interior time of our brains correspond to the interior trade time of the financial system. Statistically, the trade time is a new time scale of clock time which is transformed by multiplicative cascades into multifractal periods, according to high or low volatility. The main application of the nonlinear and multifractal approach is perhaps a better predictability. But, at least, the failures of the classical financial system can be recognized, and risks can be better calculated. Obviously, the global financial markets are governed by wild and dangerous dynamics like turbulences, hurricanes, and tsunamis in nature which must not be underestimated.

4. Complexity and Nonlinear Dynamics of Globalization and Management

Besides selection and diversity, fluctuations are needed as causes of change. In biological evolution, mutations are random events leading to new organisms that must survive the biological competition of selection. In economy, technical inventions and scientific discoveries are more or less random events leading to innovations that must become successful products in the economic competition of markets. It is obvious that diversity, self-organizing competition, and innovation are basic principles of complex dynamical systems in general. Therefore, *globalization* is not a political ideology we may like or dislike. It is a lawful, nonlinear process that we should analyze in order to handle it and to find solutions according to human interests.

Under the conditions of *nonlinear dynamics*, it is not sufficient to have good intentions, which could lead to undesired side-effects in the long run. Thus, for decisions in nonlinear dynamics, one must consider an appropriate window of time. Dynamical systems in nature and society have different scales of time. In democracies, politicians prefer to take into account short-term effects only, because they want to be elected again in the next election. Political power in democracies means power for an elected period. This is the reason why political decisions with short-term benefits may lead to disadvantages in the economic system with longer periods of causal effects. If, e.g, the central bank of a state enlarges the set of money, the first effect is an expansion of the gross national product and employment, which is followed by increasing prices and wages with a contractive effect. On the other hand, the restriction of money at first has a contractive effect on the gross national product and employment, which is later on followed by stabilized prices, wages, and an increasing gross national product. As politicians prefer the short-term benefits, there is permanent inflationary tendency endangering the value of money.

Globalization enlarges free markets from nations to the whole world. In international competition, the industry of a country is only competitive as long as the costs of production and services are not higher than the prices that can be gained for its goods on international markets. The most important part of costs is the labour costs. Therefore, the labour costs of a country must not be too high. The wages generated by international competition are called competitive wages. They are a critical control

parameter for the *international competitiveness* of a country. If competitive wages are surpassed by a country for a longer time, then the economy will decrease with high unemployment, dramatically diminished exports and enlarged imports. Finally, the wages must be decreased again. The contrary effect happens if, on the other hand, the actual wages are massively lower than the competitive wages. There is a boom with increasing exports, decreasing imports, growing profits and increasing employment. With the increasing demand for labour the wages also increase and adjust themselves to the competitive wages.

In the age of globalization there is international competition between many countries in trying to attract firms and industries from abroad. They compete with their different forms of welfare systems, taxes, and other local conditions. Sociodiversity is a feature of complex social and economic dynamics. In complex systems, nobody knows the best solutions. It is a question of learning steps and experience by trial and error. According to F. von Hayek, *competition* is a *procedure of discovering* the best solutions. In the age of globalization, the discovering process of competition is generalized for countries and nations. They emerge as global attractors in the worldwide dynamics of free markets.

On the national level, countries organize themselves with their political and economic systems. National governance is realized by national governments legitimated by democratic elections in the best case. *Governance* needs power to enforce laws, regulations, and norms. How should global governance be made possible in the age of globalization? After the Second World War the General Agreement on Tariffs and Trade (GATT) was the first step in this direction. GATT was a global, political framework of nations that wanted to take part in free markets. After 1989, GATT was improved by the World Trade Organization (WTO) with free trade of further products and services. With respect to the increasing importance of inventions and their protection by patents, the WTO was supplemented by an agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS). The WTO enforces its national members to satisfy the global rules of their agreement. Otherwise a country is excluded from the privileges of free markets by the other members. But the attraction of free markets is very strong and supports the national interest of a country. Therefore, any member tries to avoid such a situation voluntarily. In this sense the WTO has an effective system of sanctions that are applied by an international court of trade. It is remarkable that for the first time in history a political-economic system is able to enforce sanctions without military power. Military power is even excluded, because otherwise a country loses the economic benefits of the free WTO-market.

Mankind's urgent global problems like war, poverty, and ecological pollution cannot be solved, because there are no effective systems of sanctions. Good intentions are not sufficient. The WTO with its effective sanctions could be a model for successful global governance. Less developed countries have obviously less chance to compete successfully with highly developed countries on the monetary market. They need help to improve their financial systems. Nevertheless many of them can already compete with their human capital, i.e. the know-how of their people. Thus the best help is to open the markets of wealthy OECD countries for these people. There is no better way to gain the knowledge of competitors than by competing with them.

Globalization means the critical phase transition to *global governance* in the world. We need new global structures to manage the political, economic, military, and

technological power in the world according to the interests of the majority of people on earth. Global structures emerge from the nonlinear interactions of peoples, nations, and systems. At the end of the 18th century, Kant had already demanded a law of nations leading '*To Eternal Peace*' (1795). After the First World War, President W. Wilson of the United States strongly influenced the foundation of the League of Nations. After the Second World War, the United Nations (UN) presented a new chance to handle international conflicts, but they often fail because of their lack of power. The dilemma of international law is that law needs power to enforce rights and ethical norms. Therefore nations have to give up parts of their sovereignty, in order to be dominated by commonly accepted '*order parameters*'. Clearly it is a long way to global governance among autonomous nations. On the other hand, we must not forget the practical progress made by new social and humanitarian institutions of the UN. New economic, technological, and cultural networks of cooperation emerge and let people grow slowly together in spite of reactions and frictions in political reality. On the way to 'eternal peace', Kant described a federal (multi-component) community of autonomous nations self-organizing their political, economic, and cultural affairs without military conflicts. But an eminent working condition of his model is the demand that states organize their internal affairs according to the civil laws of freedom.

Are there consequences of complexity and nonlinear dynamics for *management systems* in the age of globalization? Linear decision behaviour can obviously lead to desired results only under the conditions of complete information in an environment with linear dynamics. In complex situations, agents must consider positive and negative feedback by side effects of their own nonlinear decisions and by goals and actions of other agents. Short-term thinking is dangerous in a world with delayed and long-term side effects. Learning in complex organizations means a change of mental models, strategies and decisions. It requires *nonlinear information feedback*. Thus, we should aim at improving complexity in management systems with respect to structural complexity (e.g. flat hierarchies, short length of decision processes, appropriate number of controlling units), with respect to information complexity (e.g. effective knowledge management, information symmetry and transparency), and finally with respect to the individual complexity of co-workers and the diversity of their different creative potentials.

But diversity and complexity do not automatically lead to the emergence of self-organizing fruitful effects. In firms, administrations, and other kinds of organizations, we can measure *costs of complexity* influencing their position in economic competition. Direct costs of complexity are distinguished as single direct costs for too complex products with new expensive materials, tools, and features, and permanent direct costs for service and administration of these products. Indirect costs of complexity are hidden in the organizational structure of a firm or administration. An example is the time of a manager or co-worker not being optimally used. Costs of complexity influence all steps of the value chain during an industrial production. There are, for instance, the costs of suppliers for various units of products. Furthermore, costs of complexity emerge in purchase and logistic, production and construction, marketing, and administration. If dangerous drivers and traps ("attractors") of complexity arise, the capital of a firm decreases and the energy of co-workers and managers are dissipated. Examples are offers of firms with a too much variety, over-engineering and defects in quality.

Thus, complex nonlinear organizational structures can lead to increasing butterfly effects, spreading to the whole organization (e.g. costs of complexity) like an epidemic. In complex organizations, nonlinear processes cannot be forecast in the long run. Early controlling is necessary to prevent chaotic attractors, i.e. the traps of complexity. In the competition of globalization, complex organizations only survive as learning, rapidly adapting, and flexible systems. In complex organizations, varieties of competent co-workers recognize problems better and react faster than central controlling in an over-regulated hierarchy. Therefore, we should deregulate and support *self-regulating autonomy*. Complex organizations are nonlinear social systems of people with different abilities, attitudes, emotions and interests. The sociodiversity of people is the human capital for a sustainable progress, not only in single organizations, but in the evolutionary process of globalization.

5. Complexity and Nonlinear Dynamics of Computational and Information Systems

Dynamical systems can be characterized by *information* and *computational concepts*. A dynamical system can be considered as an *information processing machine*, computing a present state as output from an initial state of input. Thus, the *computational efforts* to determine the states of a system characterize the *complexity of a dynamical system*. The *transition from regular to chaotic system*s correspond to increasing computational problems, according to increasing degrees in the *computational theory of complexity*. In statistical mechanics, the *information flow* of a dynamical system describes the intrinsic evolution of statistical correlations. In *chaotic systems* with sensitivity to the initial states, there is an increasing *loss of information* about the initial data, according to the decay of correlations between the entire past and future states of the system. In general, *dynamical systems* can be considered as *deterministic, stochastic* or *quantum computers*, computing information about present or future states from initial conditions by the corresponding dynamical equations. In the case of quantum systems, the binary concept of information is replaced by *quantum information* with superposition of binary digits. Thus *quantum information* only provides probabilistic forecasts of future states.

The idea to conceive dynamical systems as automata dates back to the mechanistic world-view of the 17th and 18th century. In the philosophy of Leibniz, even organic systems are considered as "natural automata surpassing all artificial automata infinitely". John von Neumann's concept of cellular automata gave the first hints of computational models of living organisms conceived as *self-reproducing automata a self-organizing complex systems*. The *phase space* is a homogeneous lattice which is divided into equal cells like a chess board. An *elementary cellular automaton* is a cell which can have different states, for instance, binary states of 'black' (1) and 'white' (0). An *aggregation* of elementary automata is called a composite or *complex automaton*. Each elementary automaton is characterized by its *environment*, i.e. the neighbouring cells. They change their states according to Boolean transformation rules depending on their cellular environments. The *dynamics* of a complex automaton is determined by synchronous applications of the transformation rules producing cellular patterns of black cells. These clusters can be considered as *attractors* the dynamics of cellular automata

aim at. Thus, there are classes of automata with *fixed points* and *oscillating patterns*, independent of initial cellular configurations, contrary to *chaotic patterns* of automata, depending sensitively on tiny changes of initial configurations.

Unlike Turing-like program-controlled computers, the human brain is characterized by fuzziness, incompleteness, robustness, and resistance to noise, but also by chaotic states, dependence on sensitive initial conditions, and last but not least by learning processes. These features are well known in the nonlinear complex system approach. Concerning the architecture of Turing-like computers and complex systems, an essential limitation derives from the sequential and centralized control of Turing-like computers, but nonlinear complex dynamical systems are intrinsically parallel and self-organized.

The information processing of the human brain is simulated by complex neural networks with learning algorithms. From a technical point of view, neural nets are complex systems of cells with different layers like the architecture of our cortex. Neurochemical interaction of cells is simulated by numerical weights of input data which affect firing or non-firing of technical neurons in dependence of certain threshold values. In this manner, microscopic neurons connect themselves in *macroscopic patterns*. There is no central processor or commanding neuron that can think or feel. Cognitive features of the brain are correlated to macroscopic patterns of connected neurons. Perceptions are transformed i nto neural maps o f t he b rain which c an b e c haracterized b y macroscopic *order parameters*.

Concerning the philosophical *mind-body problem*, the complex system approach states that cognitive activity is neither completely independent and different from brain activity nor simply identical, nor an epiphenomenon. Thoughts and feelings are assumed to be both product and producer of neural processes without being identical to them. In the framework of complex systems, the brain is modelled as a self-organizing system which operates far from thermal equilibrium and close to certain threshold values as instability points. During neural instability, different modes of collective excitations evolve to coherent macroscopic patterns which are neurophysiologically based on certain cell assemblies and psychologically expressed as certain feelings or thoughts.

Cognitive and *emotional* phenomena are referred to m acroscopic properties of the brain's dynamics and to order parameters of awareness which govern the underlying microscopic processes. In a situation of emotional instability, for example, a certain feeling may d ominate t he o ther o nes. T he c ompetition o f s table a nd i nstable modes i s explained by the slaving principle. In a situation of decision one thought or concept begins t o e nslave t he o ther p ossible o nes. T hese non-equilibrium p hase t ransitions a re governed by very few order parameters of preference and awareness.

Thus the complex system approach offers a research program to bridge the gap between *brain research* and *cognitive science*. In a famous metaphor, Leibniz compared the m achinery of a h uman brain and body w ith t he m achinery of a m ill w hich can be entered and observed. In modern brain research, the interacting cog wheels of the mill are the firing and non-firing neurons which could be technically constructed by a neural net. If the human brain is considered as a complex dynamical system, then emergence of mental states can be modelled by phase transitions of macroscopic order parameters which are achieved by collective nonlinear interactions of neurons, but which are not reducible to microscopic states of the system: A single neuron cannot think or feel. The complex system approach is an empirical research program which can be specified and

tested in appropriate experimental applications to understand the dynamics of the human cognitive system. Further on, it gives heuristic devices to construct artificial systems with cognitive features in robotics.

In a dramatic step, the complex systems approach has been enlarged from neural networks to *global computer networks* like the *World Wide Web*. The Internet can be considered as a *complex open computer network* of autonomous nodes (hosts, routers, gateways, etc.), self-organizing without central control mechanisms. The information traffic is constructed by information packets with source and destination addresses. *Routers* are nodes of the network determining the local path of each packet by using local routing tables with cost metrics for neighbouring routers. A router forwards each packet to a neighbouring router with lowest costs to the destination. As a router can only deal with one packet, other arriving packets at a certain time must be stored in a buffer. If more packets arrive than a buffer can store, the router discards the overflowed packets. Senders of packets wait for confirmation message from the destination host. These buffering and re-sending activities of routers can cause *congestion in the Internet*. A *control parameter* of data density is defined by the propagation of congestion from a router to neighbouring routers and dissolution of the congestion at each router. The cumulative distribution of congestion duration is an *order parameter of phase transition*. At a critical point, when the congestion propagation rate is equal to congestion dissolution, *fractal and chaotic features* can be observed in data traffic.

Congested buffers behave in surprising analogy to infected people. If a buffer is overloaded, it tries to send packets to the neighbouring routers. Therefore the congestion spreads spatially. On the other hand, routers can recover when the congestion from and to the own subnet are lower than the service rate of the router. That is not only an illustrative metaphor, but a hint on *nonlinear mathematical models* describing true *epidemic processes* like malaria extension as well as the dynamics of routers. Computer networks are *computational ecologies*. The capability to manage the complexity of modern societies depends decisively on effective communication networks.

It is not only a metaphor to transform the Internet to a *superbrain* with *self-organizing features of learning and adapting. Information retrieval* is already realized by *neural networks* adapting to the information preferences of a human user with *synaptic plasticity*. In sociobiology, we can learn from populations of ants and termites how to organize traffic and information processing by *swarm intelligence*. From a technical point of view, we need intelligent programs distributed in the nets. There are already more or less intelligent virtual organisms ('*agents*'), learning, self-organizing and adapting to our individual preferences of information, to select our e-mails, to prepare economic transactions or to defend the attacks of hostile computer viruses, like the immune system of our body.

Complexity of global networking does not only mean increasing numbers of PC's, workstations, servers, and supercomputers interacting via data traffic in the Internet. Below the complexity of a PC, cheap and smart devices of low-power are distributed in intelligent environments of our everyday world. Like GPS (Global Position System) in car traffic, things of everyday life could interact *telematically* by sensors. The real power of the concept does not come from any one of these single devices. In the sense of *complex systems*, the power emerges from the collective interaction of all of them. For instance, the optimal use of energy could be considered as a macroscopic *order parameter* of a household which is realized by the *self-organizing use* of different

household goods according to less consumption of electricity during special time periods with cheap prices. The processors, chips and displays of these smart devices do not need a user's interface like a mouse, windows, or keyboards, only just a pleasant end effective place to get things done. *Wireless computing devices* of small scales become more and more invisible to the user. Ubiquitous computing enables people to live, work, use, and enjoy things directly without being aware of their computing devices.

6. What can we learn from Nonlinear Dynamics of Complex Systems?

What are the *human perspectives* in these developments of *dynamical, information,* and *computational systems?* Modern societies, economies and information networks are highly dimensional systems with a complex nonlinear dynamics. From a methodological point of view, it is a challenge to improve and enlarge the instruments of modelling from low to high dimensional systems. Modern systems science offers an *interdisciplinary methodology* to understand typical features of self-organizing dynamics in nature and society. As nonlinear models are applied in different fields of research, we gain general insights into the predictable horizons of oscillatory chemical reactions, fluctuations of species, populations, fluid turbulence, and economic processes. Obviously, nonlinear modelling explains the difficulties of the modern Pythias and Sibyls.

The reason is that human societies are not complex systems of molecules or ants, but the result of highly *intentional acting beings* with a greater or lesser amount of free will. A particular kind of *self-fulfilling prophecy* is the Oedipus effect in which people like the legendary Greek king try, in vain, to change their future as forecasted to them. From a macroscopic viewpoint we may, of course, observe single individuals contributing with their activities to the collective macrostate of society representing cultural, political, and economic order (*order parameters*). Yet, macrostates of a society, of course, do not simply average over its parts. Its order parameters strongly influence the individuals of the society by orientating their activities and by activating or deactivating their attitudes and capabilities. This kind of feedback is typical for complex dynamical systems. If the control parameters of the environmental conditions attain certain critical values due to internal or external interactions, the macrovariables may move into an unstable domain out of which highly divergent alternative paths are possible. Tiny unpredictable microfluctuations (e.g., actions of very few influential people, scientific discoveries, new technologies) may decide which of the diverging paths in an unstable state of bifurcation society will follow.

Therefore, the paradigm of a centralized control must be given up by the insights in the self-organizing dynamics of highly dimensional systems. By detecting global trends and order parameters of complex dynamics, we have the chance of implementing favourite tendencies. By understanding complex system we can make much more progress in evaluating our information technologies and choosing our next steps. Understanding complex systems supports deciding and acting for a sustainable future of a complex world.

References

1. H.D.I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer: New York 1995
2. H. Haken/ A. Mikhailov (eds.), *Interdisciplinary Approaches to Nonlinear Complex Systems*, Springer: New York 1993
3. K. Mainzer, *Thinking in Complexity. The Computational Dynamics of Matter, Mind, and Mankind*, Springer: New York 4th edition 2004
4. K. Mainzer, *Symmetry and Complexity. The Spirit and Beauty of Nonlinear Science*, World Scientific: Singapore 2005
5. B.B. Mandelbrot, *Fractals and Scaling in Finance. Discontinuity, Concentration, Risk*, Springer: New York 1997
6. A.Scott (ed.), *Encyclopedia of Nonlinear Science*, Routledge: New York 2005
7. M. Small, *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance*, World Scientific: Singapore 2005

This page is intentionally left blank

FRACTALS, COMPLEXITY AND CHAOS
IN SUPPLY CHAIN NETWORKS

MICHAEL PEARSON

Centre for Mathematics & Statistics, School of Management,
Napier University, Craiglockhart Campus, Edinburgh EH14 1DJ

Recent work identifying an equilibrium solution in a supply chain network through cooperation and coordination along edges (links) in the network identifies the links as the stochastic modelling entities rather than the nodes of the network. This has been generalized to a supply chain network where the equilibrium solution applies to interactions between components of supply chain networks. We introduce the concept of a component decision-maker in a supply chain as a fractal. With the aid of a simple transformation we outline features of self-similarity, invariance, chaos and feedback mechanisms, and examine some repeating patterns in prediction capability control charts. We also discuss fractal geometry in the evolution of supply chain networks.

1   Introduction

B.B. Mandelbrot (1982) showed that the early mathematical 'fractals' developed by mathematicians such as Cantor (1883) and Sierpinski (1915) were far more than odd shapes deviating from familiar mathematical constructions. He demonstrated that these early fractals have many features in common with shapes found in nature. We will introduce a multi-disciplinary problem which incorporates many fractal features including self-similarity, invariance, chaos and feedback mechanisms. The problem arises in the study of supply chain networks and begins with the classical newsvendor problem. This problem addresses the sale of products with a limited selling period. It attempts to determine the optimal supply of items that minimises costs or maximises profit in the face of uncertain demand. This basic problem is known as the single echelon newsvendor problem because there is only one decision-maker. This is the newsboy who decides on how many items, Q, to order from the supplier to maximise his profit (see §2). The two echelon problem described in §2 (Lau & Lau, 2001; Silver et al., 1998) is an extension to two decision-makers such as the manufacturer and the retailer. Pearson (2003) introduced new decision variables, which enabled the reformulation of the problem by shifting the decision making from the nodes of the network structure to the edges, and hence identified an equilibrium solution in the context of a limited amount of cooperation between the operators (Carr & Lovejoy, 2000). The first variable is the mix variable, which helps to coordinate the allocation policies of the primal and dual operators. The second is the global variable, which monitors the overall increase in system output. Pearson (2004b) extends the equilibrium solution to components consisting of several operators grouped together in the supply chain (Dutta et al., 2003; Courtright et al. 1989). The equilibrium solution is then identified for the link between the primal-dual components in the same way as it is for the link between the primal-dual operators. In this paper we highlight the feature of self-similarity between operators and

135

components in supply chain networks, and introduce the invariance property required for the equilibrium solution to be unchanged when decision-making shifts from a simple primal and dual operator problem to a simple primal and dual component problem. We examine chaotic behaviour exhibited when the bullwhip effect appears in a supply chain. We discuss various feedback mechanisms, such as adaptive target control, the target gain function and prediction capability control charts (Pearson, 2000, 2004a, 2005), which are used to moderate the chaotic behaviour caused by the bullwhip effect and asymmetrical market knowledge. We identify common repeating patterns which appear in prediction capability control charts, such as the clockwise and anticlockwise patterns revealed with push-pull marketing strategies (Pearson, 2005). Fractal geometry is discussed with regard to the primal-dual decision-making elements within supply chain structures. §8 contains an illustrative example and we end with a conclusion (§9).

2   The Newsvendor Optimisation Problem

2.1   The Single Echelon Newsvendor Problem

In the single echelon problem (Silver et al., 1998) a retailer wants to find the supply of items to allocate when demand is uncertain. If F_D is the cumulative distribution function of the demand, R is the retail price, M is the manufacturing cost, while V_e is the salvage value for items returned, then the optimal order quantity, Q^* which maximises the retailer's profit is

$$Q^* = F_D^{-1}(\frac{R-M}{R-V_e}).$$ (1)

2.2   The Two Echelon Newsvendor Problem

The two echelon newsvendor problem is the problem faced by two decision-makers. One is the retailer while the other is the manufacturer. If the retailer orders too many items from the manufacturer, so that he has many unsold items at the end of the day, then he suffers a consequent loss of income due to high overage (assuming that the products are perishable). On the other hand if he orders too few items then stock-outs occur and so he suffers lost income as a result of lost opportunity (underage). He may also suffer future loss of goodwill from customers who may decide to buy elsewhere in future. The two echelons are the two decision-makers (nodes), which are the retailer and the manufacturer. The problem is known as the $\{(C,V),(Q,R)\}$ problem, since the manufacturer determines the cost to charge the retailer (C) and the credit per unit to pay for unsold units (V). On the other hand the retailer determines the amount of inventory (Q) to order from the manufacturer and the retail price (R) to charge the customer. The objectives of both the manufacturer and the retailer are to maximise profit. The stochastic problem has no generalized equilibrium solution under certain basic assumptions (Emmons & Gilbert, 1998). One assumption is that the manufacturer and retailer engage in sequential rather than simultaneous decision making. If the manufacturer pays the

retailer V for each item returned at the end of the season and the manufacturer then claims the salvage value, V_e, for items returned, the retailer's solution is to order

$$Q^* = F_D^{-1}\big[(R-C)/(R-V)\big]. \tag{2}$$

The manufacturer's optimisation problem is to find C^* that maximizes

$$\Pi_m = (C-M) \times F_D^{-1}\big[(R-C)/(R-V)\big], \tag{3}$$

where Π_m is the manufacturer's profit. The retailer is required to substitute the manufacturer's chosen value of C^* into (2) to find his optimal order decision Q^*.

2.3 Modelling Assumptions

1. We will use the single period inventory model
2. Unbiased demand and supply fitting or forecast techniques are applied
3. All prediction errors are normally distributed
4. The relationship between the standard deviation of the errors and the demand mean is well behaved, so that first and second derivatives exist and are continuous

2.4 Composing a Simple Transformation

Pearson (2003) proposed a simple transformation called the primal-dual transformation, whereby the node decision-making variables in the stochastic model are transposed into joint edge variables.

Prioritise edge over node

Figure 1. The Primal-Dual Transformation

This is illustrated in Figure 1. The primal-dual transformation prioritises edges over nodes by considering the stochastic entities defined as the sum ('global' variable) and difference ('mix' variable) of the node demand and supply variables calculated at each node associated with the corresponding edge. The primal-dual objective is to maximise

E{Profit}
   = E{Contribution from captured demand - Costs of overage - Costs of underage}
   $= 2a\mu_1 - \{2a(\phi(k)-k(1-\Phi(k))) + b(k\Phi(k)+\phi(k)) + c(\phi(k)-k(1-\Phi(k)))\}\sigma_e$ (4)

subject to:

$$\mu - k\sigma_e = 0, \quad \text{(Newsvendor Constraint)} \tag{5}$$

where a is the mean primal-dual profit per unit sold, b is the sum of the primal overage and dual underage costs and c is the sum of the primal underage and dual overage costs per unit produced. Also $E(D) + E(Q) = \eta$ and $E(Q) - E(D) = \mu$ and σ_e (the standard deviation of the forecasting errors) depends on μ and η.

2.5 *Primal-Dual Solution for Variability dependent on* μ *and* η

We solve the stochastic equations (4) and (5) using a Lagrange multiplier (Pearson 2003), where the independent variables are k, μ and η. The solution is:

$$\Phi(k) + \phi(k)\sigma_e^\mu = \frac{a+c}{b+c} \tag{6}$$

$$\phi(k)\sigma_e^\eta = \frac{a}{b+c} \tag{7}$$

where $\sigma_e^\mu, \sigma_e^\eta$: stand for $\partial\sigma_e / \partial\mu$ and $\partial\sigma_e / \partial\eta$ respectively. Also $\phi(\bullet)$ and $\Phi(\bullet)$ are the standard normal distribution density function and cumulative distribution function respectively. The assumption A5 helps to maintain the independence between the mix and global variables Q-D and Q+D (and so also μ and η). As a result both the simplicity and the complexity of the solution can be displayed more effectively. Figures 4 and 5 illustrate the distinct behaviour displayed by the error standard deviation, σ_e when it is plotted against the mix variable, μ and the global variable, η.

3   Extension to the Complete Network

3.1 *Supply Chain Entities and Links*

Figure 2 illustrates an operator in the supply chain network. If he looks downstream then he acts as a dual operator who has a link to a downstream primal operator. If he looks upstream then he acts as a primal operator with a link to an upstream dual operator. We employ the transformed link stochastic variables $(k, \mu, \eta)^d$ for the downstream link and the variables $(k, \mu, \eta)^u$ for the upstream link. The solution sets associated with the downstream and upstream variables are $S(k, \mu, \eta)^d$ and $S(k, \mu, \eta)^u$, respectively.
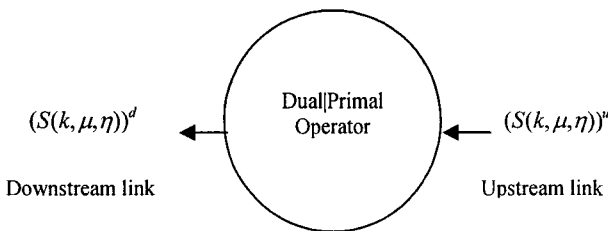
Figure 2. Dual|Primal Supply Chain Operator

The simple structure illustrated in Figure 2 can be extended to a tree structure with a simple operator and more than one upstream or downstream links (Pearson, 2004b). The combined upstream flow would then be modelled as the upstream flow and the combined downstream flow as the downstream flow.

4   Self Similarity

We can treat sets of operators as a component. The component then has a total upstream and downstream flow and is related to other operators or components in the same way that operators are related to each other. The component set is then a collection of copies of the operator set with exactly similar dual|primal relations.

4.1 Chain Structure

Figure 3 illustrates self-similarity in a simple supply chain network. The dual|primal component is a collage of two similar copies of the dual|primal operator. The solution set $S(k,\mu,\eta)$ is associated with the link between the decision-making operators who have cooperated by prioritising their edge (link) variables (k,μ,η) over their node variables ($D,\sigma_{e_D},Q,\sigma_{e_Q}$) and agreed to the coordination (newsvendor) constraint, $\mu - k\sigma_e = 0$ where σ_{e_D}, σ_{e_Q} and σ_e are the standard deviations of the primal, dual and joint prediction errors, respectively. When decisions between components are being addressed the solution sets are $(S(k,\mu,\eta))^d$ and $(S(k,\mu,\eta))^u$ for the downstream and upstream links respectively. Decisions between components reduce the decision-making capacity of the operators within the components.
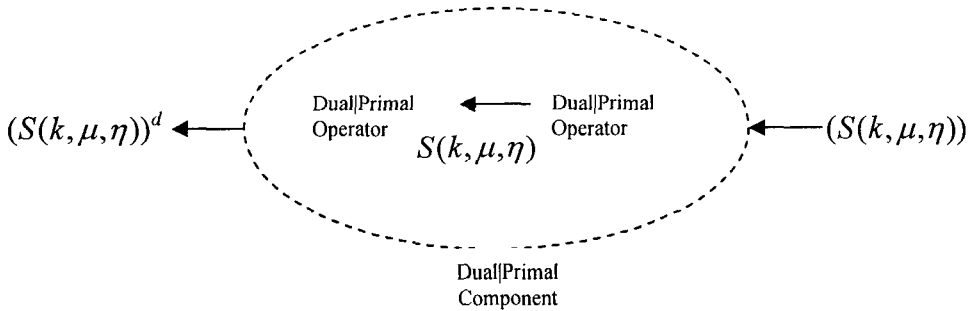
Figure 3. Self-Similarity in Simple Chain Structure ($S(k,\mu,\eta)$)

4.2 Invariance Property

We examine conditions where the solution is unchanged by the switch from operator to component level decision-making. We have that the expected demand and supply at the operator level within the Dual|Primal component illustrated in Figure 1 are:

$$E(D) = \mu_D = ((\mu_D + \mu_Q) - (\mu_Q - \mu_D))/2 = (\eta - \mu)/2$$

$$E(Q) = \mu_Q = ((\mu_D + \mu_Q) + (\mu_Q - \mu_D))/2 = (\eta + \mu)/2 \,.$$

Similarly the expected component upstream demand and downstream supply are

$$E(D)^u = \mu_D^u = (\eta^u - \mu^u)/2$$

$$E(Q)^d = \mu_Q^d = (\eta^d + \mu^d)/2 \,.$$

Now if the solution is unchanged then

$$E(D) = E(Q)^d \quad \text{and} \quad E(Q) = E(D)^u$$

So that
$$\eta - \mu = \eta^d + \mu^d \qquad\qquad (8)$$

$$\eta + \mu = \eta^u - \mu^u \,. \qquad\qquad (9)$$

Subtracting (8) from (9) and adding (8) to (9) gives

$$\mu = \left((\eta^u - \eta^d) - (\mu^u + \mu^d)\right)/2 \qquad\qquad (10)$$

$$\eta = \left((\eta^u + \eta^d) - (\mu^u - \mu^d)\right)/2 \qquad\qquad (11)$$

This result could occur when the operators have symmetrical market knowledge, so that the operators know as much about demand and supply throughout the chain as the component decision-makers do. We generally seek conditions based on increased channel optimality and capability (Pearson, 2004b) before handing over decision-making from operator level to component level, and so the invariance property is not a sufficient incentive to change to component level.

5   Chaos

5.1 Bullwhip Effect

The bullwhip effect is a feature of supply chains which up till now has defied simple modelling. It describes the counter-intuitive situation whereby orders at the higher (upstream) levels of the supply chain can exhibit more variability than those at customer (downstream) level. 'Demand signal processing', the 'rationing game', 'order batching', and 'manufacturer price variations' all increase the bullwhip effect (Silver et al, 1998). The effect is also increased if demand variability changes, which is often the case when new markets are sought.

5.2 Asymmetric Market Knowledge

Pearson (2003) describes the chaotic behaviour in a supply chain when there is an increase in both primal and dual variability. Each operator seeks to maintain his level of profit achieved during stable operation at the expense of the other operator, which leads to disequilibrium when market knowledge is asymmetric. In practice the operator with

stable variability penalizes the operator carrying the increased variability. Since the downstream operator (retailer) often carries the burden of increased variability, improved efficiency tends to pass upstream towards the dual operator (manufacturer), not downstream towards the primal operator (retailer). The primal-dual formulation provides an equilibrium solution to this complex, asymmetric multi-echelon problem. The operators cooperate through limited information exchange and then coordinate their allocation strategy.

6 Feedback Mechanisms

Feedback mechanisms are a feature of industrial and business processes (Box & Lucerno, 1997). In the context of fractal problems their establishment counters chaotic behaviour. Pearson (2000, 2004a, 2005) proposes three such mechanisms.

6.1 Adaptive Target Control

The method of adaptive target control acts as a feedback system which enables appropriate capable targets to be selected when a system is not capable of achieving the agreed or imposed target values. The targets refer to levels of overage and underage which occur when produce is distributed in the supply chain. If a forecasting technique is unable to achieve both targets simultaneously (which is what is meant by saying that the prediction method is capable of achieving both targets, so that the prediction capability, $C_{PR} = 1$) then a theorem identifies adapted targets, which the prediction method is capable of achieving.

6.2 Target Gain Function

The target gain function is a feedback mechanism that addresses the problem of accurately identifying uncertainty in the measurement of the standard deviation of the errors. An interval estimate making use of a target gain function to identify improved solutions to the newsvendor problem is introduced. Current practice targets for overage and underage are used as the starting point for the identification of improved targets.

6.3 Primal-Dual Capability Charts

The prediction capability chart (Pearson 2003, 2005) monitors changes in forecast variability as the mix variable changes. Figure 4 illustrates the chart and shows that the region of improved optimality associated with a given solution, $(\bar{x}, s)$ bounded by the isovalue line and the μ axis, contains within it the region of improved capability. The control zone is the area of stable cooperation between the primal and dual operators. It can be calculated dynamically from a time series by using a confidence interval $[\bar{x} \pm t_{df, \alpha/2} s]$ for the mix variable, μ and a χ^2 interval $[s_1, s_2]$ for σ_e using the point estimate, s, calculated at the optimal solution. The control zone shows when the allocation method is out of control. The chart itself acts as a feedback mechanism for the decision-maker when allocating inventory.
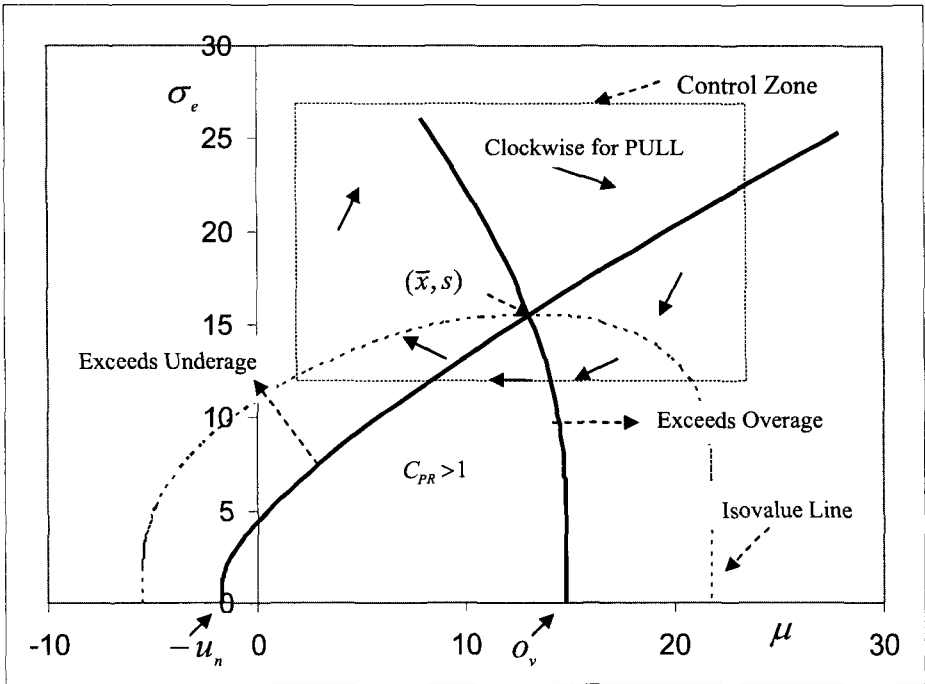
Figure 4. Prediction Capability Control Chart

6.4 Repeating Push-Pull Patterns

Out of control points could be corrected by use of equations (6) and (7) (or the minimisation of costs solution outlined in Pearson 2003) to identify a new value of k which brings the process back into control. The push-pull marketing strategies are identified on the prediction capability control plots and charts (Pearson, 2005). The cycle of points in the prediction capability chart may follow an anti-clockwise pattern, which is typical of the push strategy employed in marketing (Cachon, 2004). This contrasts with a clockwise pattern which shows that customer demand for the product generated the increase in supply, which tends to happen with the pull strategy. In contrast to the prediction capability chart for the mix variable is the primal-dual global plot for the data.

The global plot of σ_e against η often has a positive slope indicating that the error variability increases for higher output levels in the global market. Figure 5 shows the efficient frontier (represented by the isovalue line in Figure 4) advancing as output increases. This is prevalent in a market where the actions of a competitor may cause variability in customer demand which is difficult to predict.
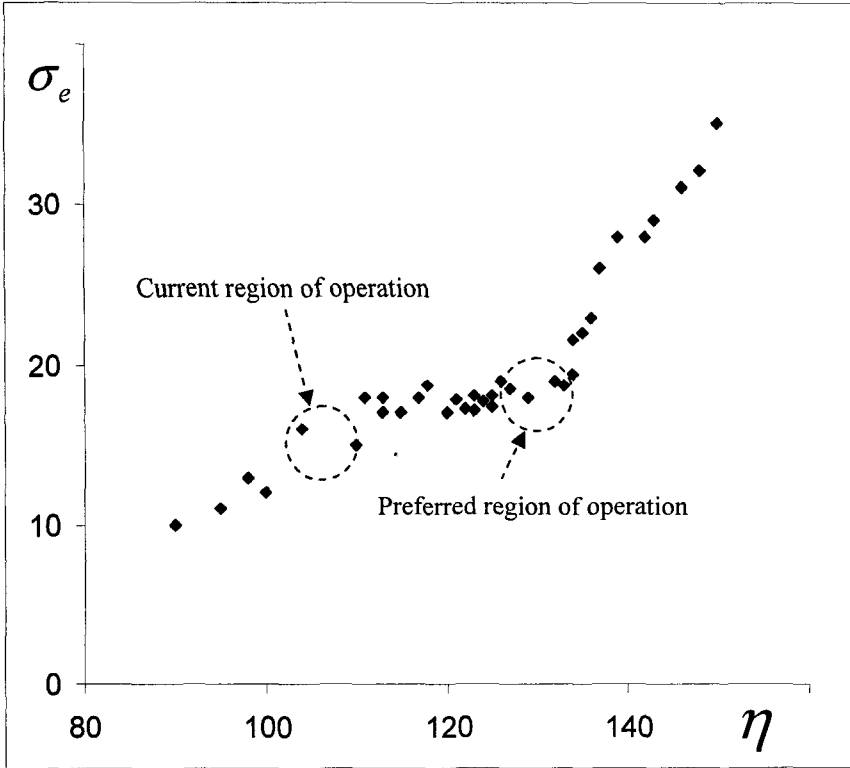
Figure 5. Primal-Dual Global Plot

7 Fractal Geometry

We break down the geometrical distribution problem into decision making components. We formulate the optimisation problem of each component. We apply the primal-dual transformation to the links between each component (prioritise edge over node) and reformulate the optimisation problem as a primal-dual problem. We then interrogate whether the problem has symmetrical or asymmetrical market knowledge and so establish whether the invariance property applies. We investigate the bullwhip effect and chaotic behaviour. We establish feedback mechanisms. We test for the principles of optimality and capability. We then investigate the fractal geometry of the problem, as reformulated into new decision-making components.

8 Illustrative Example

We consider a manufacturer and retailer who jointly decide which area of the market to operate in and the marketing strategy they intend to pursue. Using data drawn from recent historical demand and supply patterns they formulate primal-dual capability mix and

global plots. Their preferred marketing strategy is the pull mechanism based on maintaining a solid customer base. Their strategy is confirmed by:

1. the clockwise pattern of the capability chart (Figure 4), which shows that surges in demand precede the matched supply figures.

2. the in-control pattern of points on the capability chart (Figure 4). This ensures capable operation and optimisation while staying within target margins. This ensures viable business operation while maintaining customer satisfaction and goodwill.

3. the global plot operating at maximum volume output within a low variability window of operation. In Figure 5 the current area of operation is set at $\eta = 105$ and $\sigma_e = 15$ (refer to Figure 4, optimal solution at $\mu = 12$ and $\sigma_e = 15$). The proposal is to increase volume to the preferred region of operation where $\eta = 125$ and $\sigma_e = 18$. This is at the cusp of the efficient frontier before possible chaotic behaviour begins.

4. the fact that occasional promotional activity which extends operation into the area of increased variability (identified from the high volume output, $\eta > 125$, in Figure 5) gives an anti-clockwise pattern on the capability plot (Figure 4), which indicates a change to a push strategy.

5. the fact that the push strategy may lead to out of control operation beyond the extreme of the efficient frontier. Factors such as increased competition and the breakdown of coordinated activity within the supply chain become significant.

The manufacturer (or retailer) may attempt to gain ownership of the supply chain. In this case decision-making would shift to the component level consisting of the retailer-manufacturer component (Figure 3). The self similar principles would apply to decision-making at this component level as at the operator level. The owner of the supply chain would want, at least, to maintain invariance within the chain, but would hope to increase overall capability and optimality (Pearson, 2004b).

9 Conclusion

We have identified a fractal (primal|dual) element as a decision-maker in a supply chain network who is related upstream with a dual element and/or downstream with a primal element. The decision maker can be a single operator in the supply chain, such as a manufacturer or distributor, or a group of operators considered as a decision-making element called a component. The operators and components display fractal features such as self-similarity, invariance (in the presence of the equilibrium solution), chaos, feedback mechanisms and fractal geometry in the context of complex supply chain network structures. The complexity of such real-life systems is realised within a relatively simple modelling framework. Equilibrium, order and harmony are starkly contrasted with chaos and the search for corrective feedback mechanisms. We are able to realistically model such phenomena as the bullwhip effect and asymmetric market knowledge and effectively remedy some of the consequent excesses by using appropriate feedback mechanisms.

Acknowledgements

References

1. Box, B., A. Lucerno. 1997. Statistical Control by Monitoring and Feedback Adjustment. John Wiley & Sons.
2. Cachon, G.P. 2004. The Allocation of Inventory Risk in a Supply Chain: Push, Pull, and Advance- Purchase Discount Contracts. *Management Science*. 50, (2), 222-238.
3. Cantor, G. (1883) *Uber unendliche, lineare Punktmannigfaltigkeiten V,* Mathematische Annalen **21** 545-591.
4. Carr, S., W. Lovejoy. 2000. The Inverse Newsvendor Problem: Choosing an Optimal Demand Portfolio for Capacitated Resources. *Management Science*. **46**(7), 912-927.
5. Courtright, J., Fairhurst, G. and Rogers, L. 1989. Interaction Patterns in Organic and Mechanistic Systems. *Academy of Management Journal*. **32(4)**: 773-802.
6. Dutta Bhaskar, S. Ghosal and D. Ray. 2003. Farsighted Network Formation. *Journ. of Econ. Theory* (to appear).
7. Emmons, H., Gilbert, S.M., 1998. The Role of Returns Policies in Pricing and Inventory Decisions for Catalogue Goods. *Management Science*. 44, 276-283.
8. Lau, A.H., H.S. Lau. 2001. Some Two-Echelon Style-Goods Inventory Models with Asymmetric Market Information. *Europ.Journ.of Opl.Res*. **134**(1), 29-42.
9. Mandelbrot, B.B. 1982. The Fractal Geometry of Nature. Freeman, New York.
10. Pearson, M.A., 2000. The Incorporation of Target Performance Measures and Constrained Optimisation in the Newsboy Problem. *Journ. of Oper. Res. Soc.* **51**(6) 744-754.
11. Pearson, M.A. 2003. An Equilibrium Solution to Supply Chain Synchronisation. *IMA Journal of Management Mathematics*. **14** 165-185.
12. Pearson, M.A. 2004a. The Application of Prediction Capability and Adaptive Target Control to the Newspaper Business. *Europ. Journ. of Oper. Res.* (In Press. Available electronically, August 2004).
13. Pearson, M.A. 2004b 'Decision-Making in Supply Chain Networks' *Proceedings of MUDSM conference*, Coimbra, Portugal (Sept. 2004).
14. Pearson, M.A. 2005. Prioritising Edge over Node: Process Control in Supply Chain Networks. *Proceedings of IMSIO Conference*, Salford, UK (June, 2005).
15. Sierpinski, W. 1915. *Sur un courbe cantorienne don't tout point est un point de ramification.*C.R. Acad. Paris 160 302.
16. Silver, E.A., F.P. Pyke, R. Peterson. 1998. Inventory Management and Production Planning and Scheduling. John Wiley & Sons.

This page is intentionally left blank

THE EFFECTS OF DIFFERENT COMPETITION RULES ON THE POWER-LAW EXPONENT OF HIGH INCOME DISTRIBUTIONS

KEIZO YAMAMOTO

*Faculty of Engineering, Setsunan University,17-8 Ikeda-Nakamachi,
Neyagawa, Osaka 572-8508, Japan. Email: yamamoto@ise.setsunan.ac.jp*

SASUKE MIYAZIMA

*Department of Natural Science, Chubu University,1200 Matsumoto,
Kasugai, Aichi 487-8501, Japan. Email: miyazima@isc.chubu.ac.jp*

HIROSHI YAMAMOTO, TOSHIYA OHTSUKI, AKIHIRO FUJIHARA
*Graduate School of Integrated Science, Yokohama City University, 22-2 Kanazawaku-seto,
Yokohama, Kanagawa 236-0027, Japan. Email: hirobrd@abox.so-net.ne.jp,
ohtsuki@yokohama-cu.ac.jp, fujihara@yokohama-cu.ac.jp*

A model for power-law problem in high-income distribution is proposed. This model is so simple that we can examine the effect of some differences in rules that allocate different amounts of resources to winners and losers in economic competitions where we obtain different power-law exponents as well as exact analysis by the master equation.

Keywords : high income model, power-law distribution, fractals, master equation

1 Introduction

Economic activities have a very intricate structure which is influenced by many factors such as production of natural resources, disasters, labor power supply, relation between governments, and so on. The economic system can be viewed as a prototype of complex systems. The gains which are mainly obtained after competitions among companies are distributed to employees. These processes perform endlessly complicated nests of trades, producing incomes from unspecific random objects. Therefore we believe that there is an intricate nest of structures and we would like to show an existence of some kind of fractality.

Power-law distributions for high-income have been investigated since 1897 [1 - 8]. Other power-law distributions are found in social and economical events such as word counting [9], family name distribution [10], firm size [11], fluctuation in finance [12], number of passengers at stations [13] as well as in high-income ranking problems. In Table 1 we show some power-law exponents of high-income data. In high-income ranking problem, it is our main interest to know the reason why the high-income distribution shows power-law and what kinds of factors determine the power-law exponent. In addition, it is also interesting to investigate what the power-law exponent means in the real economical phenomena. In this paper we suggest a model which indicates some different power-law exponents depending on competition rules.

147

When we change the competition rule in our model, different power-law exponents are obtained. When we can investigate and analyze the relation between the competition rule and the power-law exponent sufficiently, it will become clear what is a main competition rule that controls our real economics. Our model is simple enough to get an exact solution by an analytical method.

Table 1. Some earlier data of the power-law exponent (X: Income, $P(t>X)$: The probability with income more than X, R: The rank means the number of persons who earn more than X. Equation $R \sim P(t>X) \sim X^a$)

Year	Name	α	Income Data
1897	Pareto	1.5	Some Euro cities
1953	Champernone	1.7	U.K. 1951/1952
1959	Lydall	1.5	U.K. 1954/1955
1983	Montroll, et al.	1.63	USA 1935/1936
1998	Takayasu, et al.	0.5	Model
2000	Aoyama, et al.	2.05	Japan 1998
2002	Yamamoto, et al.	1.39	Model
2002	Yamamoto, et al.	1.32	Japan & USA 1998
2002	Kawamura, et al.	1.0	Model

2 Model and Simulation

A present model is constructed based on real economic activities. Companies, shops and people are simply called a member. Each member works to earn money. We assume that our system consists of N homogeneous members. The number of members N is fixed. As a basic procedure we have the following two steps, which will be modified as given by Type 1, Type 2 and Type 3. Each modification will induce a different power in the power-law.

Step1. One of two members, who are chosen randomly from the members, gets their total money after the following competition rule.

Step2. The total resource money S is kept constant during the process. Each member keeps one unit of money as a minimum amount. S is fixed except for the Type 2.

Type 1 [14]: One of the two members collects all the money of the two members, and the other loses all of one's resource money. In order to maintain the number of active members with resource money, we add one unit of money to the loser. In order to keep the whole resource money in our system we reduce one unit from a randomly selected member who has the resource money more than or equal to two units. Here, the total amount of the resource money S is fixed at 2N.

Type 2: One of the two members acquires all the money of the two members, and the other loses all of one's resource money. In order to maintain the number of active members with resource money, we add one unit of money to the loser. Therefore, the total

resource money increases by one unit per competition. We relax the above basic condition Step2 and the system is simulated under the condition $N^2 < S < (N+1)^2$ for the member N and the total resource money S.

Type 3: When amounts of the resource money for the two members are different, a winner acquires less amount of two resource moneys and the other loses the same amount of money. A winner who has smaller amount of money doesn't get all of the money of the loser. The remaining part of the rule is almost the same as for Type 1. Here, the total resource money S is fixed at $7N$.
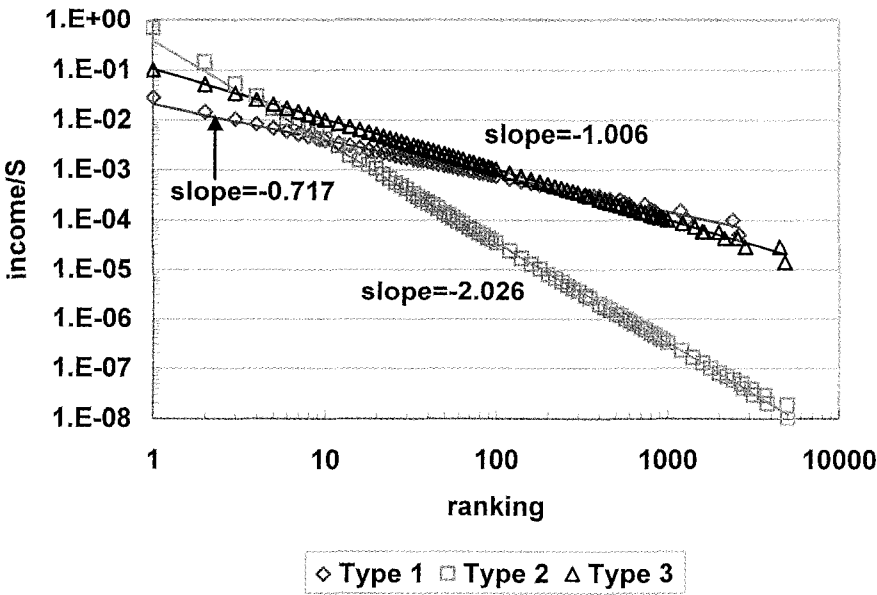
Fig. 1. Three log-log plots of income against ranking, $N = 10,000$.

Here the process which consists of a Step1 and Step2 is called as one Monte Carlo step, hereafter. Some simulation results of three different types against ranking are shown in Fig. 1. The power-law exponent of incomes against ranking of Type 1, Type 2 and Type 3 are −0.717, −2.026 and −1.006, respectively. Simulation results of the frequency F(X) against the income X are shown in Fig.2 for three different types. The power-law exponent of the frequency F(X) against the income X of Type 1, Type 2 and Type 3 are −2.45, −1.55 and −1.98, respectively. Now we compare above exponents with α in Table 1. We calculate the exponents of the cumulative quantity of Type 1, Type 2 and Type 3, −1.45, −0.55 and −0.98, respectively from the exponents of the frequency F(X) in Fig. 2. The exponent of Type 1 corresponds to the indices of α of Pareto's, Champernowne's, Lydall's, Montroll's and Yamamoto's data. The Type 2 corresponds to the index of α of Takayasu's data. The Type 3 corresponds to the index of α of Kawamura's data.
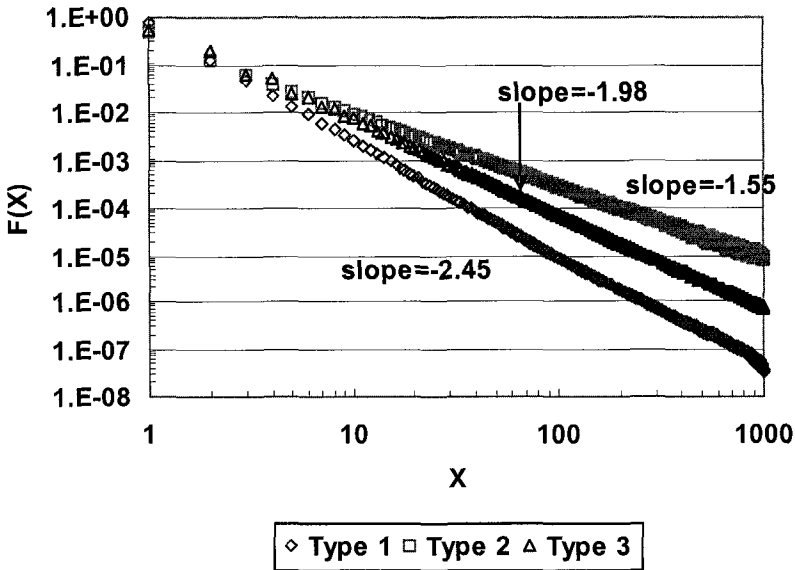
Fig. 2. Tree log-log plots of the frequency $F(X)$ against the income X, $N = 10,000$.

3 Mathematical Analysis

We introduce a probability $P(X)$ that an arbitrary chosen member holds X units of re-source money. The master equation is

$$dP(X)/dt = (1 - P(1))(1 - P(1)) + P(2) / (1 - P(1)) - P(1)^2 \qquad \text{for } X=1 \qquad (1)$$

and

$$dP(X)/dt = \sum_{i+j=X} P(i)P(j) + P(X+1) / (1 - P(1)) - 2P(X) - P(X) / (1 - P(1))$$

$$\text{for } X \geq 2, \qquad (2)$$

where the first and the second terms on the right hand side in Eq.(1) and Eq.(2) are in-crements of P(1) and P(X) for $X \geq 2$ respectively. The third term on the right hand side in Eq.(1) and the third and the fourth terms on the right hand side in Eq.(2) are decrements of P(1) and P(X) for $X \geq 2$ respectively. In Eq.(1), the first term on the right hand side is the transition probability to P(1) from all other states except P(1). The second term on the right hand side is the transition probability from P(2) to P(1) (by reducing one unit from the member in order to keep the total resource money S). The third term is the transition

probability from P(1) to P(2). In Eq.(2), the first term on the right hand side is the sum of the incoming transition probability to P(X). The second term on the right hand side is the transition probability from P(X+1) to P(X) (by reducing one unit from the member in order to keep the total resource money S). The third term on the right hand side is the sum of the outgoing transition probability from P(X). The fourth term on the right hand side is the transition probability from P(X) to P(X-1) (by reducing one unit from the member in order to keep the total resource money S). The similar equation can be seen in the cluster-cluster aggregation model and the Smoluchowski equation [15]. But the exact solution for such a model is quite difficult and sometimes impossible to obtain. If the sticking coefficients are constant for any pair of clusters and the breakup rates are specific value in the aggregation model, the generalized equation [16] corresponds to our Eq.(2).

Let us now consider the stationary state in which the probability P(X) is time-independent. If (per one Monte Carlo step) the same number of transitions occur in the incoming and the outgoing direction, P(X) does not change. This requirement can be written mathematically in the following set of equations for Type 1 model.

$$(1 - P(1))(1 - P(1)) + P(2) / (1 - P(1)) = P(1)^2 \qquad \text{for} \quad X=1 \qquad (3)$$

and

$$\sum_{i+j=X} P(i)P(j) + P(X+1) / (1 - P(1)) = 2P(X) + P(X) / (1 - P(1)) \quad \text{for } X \geq 2 . \qquad (4)$$

We define a generating function G(z) by

$$G(z) = \sum_{X=1}^{\infty} P(X) z^X \qquad |z| \leq 1, \qquad (5)$$

then the generating function G(z) of Eq.(3) and Eq.(4) becomes as follows;

$$(1 - P(1)) G(z)^2 + (z^{-1} - 3 + 2P(1)) G(z) - P(1) + z = 0 . \qquad (6)$$

The first derivative of Eq.(6) becomes a trivial equation where z=1. Now if we take the 2nd derivative of Eq.(6) and let z=1, the quadratic equation of the first moment G'(1) is obtained

$$(1 - P(1)) G'(1)^2 - G'(1) + 1 = 0. \qquad (7)$$

As the first moment means the expectation, the Eq.(7) has to one real solution. Then we can derive the values for G'(1) and P(1);

$$P(1) = 3/4 \quad \text{and} \quad G'(1) = <X> = 2 \ .$$

When we substitute the value $P(1)$ into Eq.(6), the solution for $G(z)$ becomes as follows;

$$G(z) = 3 - 2z^{-1} + 2Z^{-1}(1 - z)^{3/2} \ . \tag{8}$$

We immediately obtain the probability $P(X)$ with the aid of the Taylor expansion of the function $G(z)$ by taking the X's derivative of $G(z)$ and dividing it by $X!$.

$$P(X) = 6 (2X - 2)! / ((X + 1)!(X - 1)!2^{2X}) \ . \tag{9}$$

If we use Stirling's formula, the probability $P(X)$ shows the power-law behavior as follows;

$$P(X) \sim X^{-5/2} \ . \tag{10}$$

Therefore the cumulative distribution function is $P(>X) \sim X^{-3/2}$. This power-law index shows good fit of our simulation in Fig.2 and real data on Table 1. We can also derive the power-law index of Type 2 model similarly.

References

1. V. Pareto, *Le Cours d'Economie Politique*, Macmillan, London (1897).
2. G. D. Champernowne, *Economic Journal*, **63**, 318 (1953).
3. H. F. Lydall, *Econometrica*, **27**, 110 (1959).
4. M. F. Shlesinger and E. W. Montroll, *Lecture Note in Math*, **1035**, 138 (1983).
5. H. Takayasu and K. Okuyama, *Fractals*, **6**, 67 (1998).
6. H. Aoyama, W. Souma, Y. Nagahara, M. P. Okazaki, H. Takayasu and M. Takayasu, *Fractals*, **8**, 293 (2000).
7. K. Yamamoto and S, Miyazima, *Sociological Theory and Methods*, **16**, 133 (2001) (In Japanese).
8. K. Kawamura and N. Hatano, *J Phys Soc Japan*, **71**, 1211 (2002).
9. G. K. Zipf, *Selected studies of the principle of relative frequency in language*, Harvard University Pree, Cambridge, MA (1932).
10. S. Miyazima, Y. Lee, T. Nagamine and H. Miyajima, *Physica A* **278**, 282 (2000).
11. M. H. R. Stanley, L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhron, P. Maass, M. A. Salinger and H. E. Stanley, *Nature*, **379**, 804 (1996).
12. B. B. Mandelbrot *Fractals and Scaling in Finance in France*, Springer, Berlin (1997).
13. H. Fujita and S. Miyazima, *Master Thesis of H. Fujita,* Department of Engineering Physics, Chubu University, March (2003).

14. K. Yamamoto, S Miyazima, K.R. Koshal and Y. Yamada, *Physica A* **314**, 743 (2002).
15. R. Badii and A. Politi, *Complexity – Hierarchical Structures and Scaling in Physics*, Cambridge University Press, Cambridge (1997).
16. T. Vicsek, *Fractal Growth Phenomena*, World Scientific, Singapore (1989).

This page is intentionally left blank

SYNERGETICS ON ITS WAY TO THE LIFE SCIENCES

HERMANN HAKEN

Institute for Theoretical Physics I, Center of Synergetic University of Stuttgart, Pfaffenwaldring 57
70550 Stuttgart, Germany

After a brief description of the basic concepts of synergetics in a qualitative and quantitative manner I recall the goals of phenomenological synergetics, semantic synergetics and microscopic synergetics. The paper focuses on explicit applications of semantic and phenomenological synergetics to psychophysics, decision making, management, and health care.

1 Synergetics: Its General Frame

1.1. *Goal*

Synergetics is an interdisciplinary field of research that deals with complex systems [1]. These are systems that are in general composed of many components interacting with each other. The interactions and/or components and/or behavior may be complicated. In many cases these systems can be considered at the microscopic and the macroscopic level. At the microscopic level we are dealing with the individual parts of a system and their interactions. A few examples may illustrate the point: in physics the elements may be molecules in liquids or gases or electrons in semiconductors or light emitting atoms in lasers. In chemistry they are molecules undergoing chemical reactions and perhaps also diffusion. In biology they may be cells in an organism or individual animals in species or neurons in a brain. In sociology they may be individuals in a society or in economy individual companies. These examples do not exhaust the field of synergetics at all.

In many - if not all - cases just mentioned the individual parts are able to form spatiotemporal or functional structures at macroscopic scales, e.g. the molecules in a liquid may form specific macroscopic movement patterns of the whole fluid, or people in a society may form specific public opinions or social groups, or the neurons in a brain may cooperate so as to produce speech or to enable us for pattern recognition. Synergetics deals with those systems that are able to form their macroscopic structures by means of self-organization, i.e. there is no specific external force necessary to impose the structures on the systems, rather they obtain these structures by their own action.

Synergetics aims at reducing the complexity of a system and as it turns out synergetics shows how to control a system indirectly rather than by steering its individual parts directly. The central question synergetics asks is "Are there general principles for self-organization irrespective of the nature of the individual parts?" At first sight such a question may seem absurd because of the quite different nature of the individual parts that must be taken into consideration. In spite of this, however, the question can be answered

in the positive at least for a large range of phenomena provided we focus our attention on those situations where the macroscopic behavior of a system changes qualitatively. What is meant by this can be best illustrated by an example taken from ordinary life. While water is composed of H_2O molecules it can appear in quite different macroscopic states, namely liquid water, solid ice or vapor. Consider the transition from water to ice. While the molecules by themselves do not change in these two states of aggregation, the macroscopic qualities are quite different, e.g. ice has different mechanical, optical, and thermal properties than water. Such transitions and similar ones in systems in thermal equilibrium have been treated in physics to a large extent both theoretically and physically and there is a sizable literature in this field.

Synergetics on the other hand goes far beyond these phenomena of so-called phase transitions because it deals with systems away from equilibrium or with non-physical systems which share, however, the same property, namely changing their macroscopic properties qualitatively.

In order to be more explicit I will first discuss basic concepts of synergetics in a qualitative way and then briefly indicate a quantitative approach based on mathematics.

1.2. Basic Concepts

1.2.1. Qualitative aspects

We consider as a system an ensemble of elements of which each has well defined specific properties and the elements interact with each other. The system is assumed to have a well defined boundary that separates it from its surroundings. The system may be influenced either from the outside or by itself by means of control parameters, e.g. in physics these may be the energy influx into a liquid or into a laser. In brains these may be neurotransmitters or hormones, etc. In the theoretical approach it is assumed that these control parameters are kept time-independent, at least for some interval of time. When control parameters are changed two situations may occur: the system either adapts smoothly or it changes its behavior qualitatively. Such a latter situation is called instability. Synergetics studies system close to such instabilities whereby it turns out to be possible to extrapolate from this situation into the newly developing stable regime.

The central insights gained by synergetics are as follows: close to instability points a new set of variables, describing the macroscopic behavior, evolves. These are the order parameters. They are much less numerous than the originally present individual parts. The order parameters determine the behavior of the individual parts via the slaving principle. This can be visualized by comparing the order parameters with puppeteers that let the individual puppets dance. But in contrast to the action of a puppeteer and the puppets following his or her actions, the individual parts can act on the order parameters. By their collective action they generate and maintain the order parameters. This leads to the principle of circular causality in which the order parameters determine the behavior of the

individual parts which in turn determine the behavior of the order parameters. When we go around the circle starting from the individual parts, it appears as if the individual parts have found their collective action and thus they are organizing themselves. This is a phenomenon of self-organization. On the other hand when we start from the order parameters and return to them, the behavior of the system is determined and described by the - in general few - order parameters with their low dimensional dynamics. This leads to an enormous information compression.

Order parameters and slaving principle allow us to make contact with chaos theory and the theory of fractals which are closely related to each other. As is well known, chaos theory studies the complex dynamics of few degrees of freedom. Nevertheless, chaos theory has been applied to systems with many degrees of freedom under the assumption that those are describable by few degrees of freedom. Quite clearly, our theory on order parameters and the slaving principle provides us with a firm basis of the just mentioned assumptions.

1.2.2. Basic concepts: Quantitative

We consider the system as a *dynamical system* that is described by its state vector $\underset{\sim}{q}(t)$

$$\underset{\sim}{q}(t) = (q_1, q_2, ..., q_n) \tag{1}$$

where $q_1, q_2, ..., q_n$ are variables that describe the time-dependence of the states of the individual parts. Note that these q_j may be vectors by themselves. In the spirit of dynamical systems theory we assume that the evolution in time of the state vector can be described by evolution equations that we assume in the general form

$$\frac{d\underset{\sim}{q}}{dt} = N\left(\underset{\sim}{q}, \alpha\right) + \underset{\sim}{F}(t). \tag{2}$$

In it N is a nonlinear function of the state vector at time t and N depends also on a set of fixed control parameters α. $F(t)$ represents stochastic forces, i.e. chance events acting on the system. Also more general forms of F in which case F depends on the state vector $\underset{\sim}{q}$ are possible (Ito-Stratonovich calculus). We assume that for a fixed set of control parameter values, a solution q to (2) is known. We then study the stability of this solution when the control parameter is changed to a new value which at least in general is done by linear stability analysis.

$$\underset{\sim}{q_0}(t) \rightarrow \underset{\sim}{q_0} + \underset{\sim}{v}(t) \tag{3}$$

This analysis reveals that close to instability points new collective variables can be defined, namely the order parameters $\underset{\sim}{\xi}$. As the detailed theory shows, the order parameters obey equations of the form

$$\frac{d\xi}{dt} = N_r\left(\xi, \alpha\right) + F_r(t) .$$ (4)

that describe a low dimensional dynamics. In contrast to conventional dynamic system theory, our approach takes into account that the fluctuating forces F are present and play just close to instability points a crucial role. The slaving principle states that the state vector q can be determined by a specific function f

$$q(t) = f\left(\xi(t), t\right)$$ (5)

where the order parameters and the state q are taken at the same time whereas there is still an explicit dependence of f on t because of the fluctuating forces. Eq. (4) expresses the result of information compression. The complex system is described by low dimensional dynamics of its order parameters. According to the slaving principle the
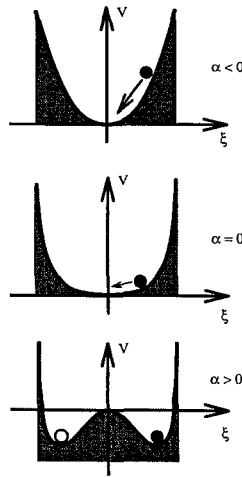
Figure 1. Visualization of the behavior of a single order parameter in the case of (non-equilibrium) phase transition. The size of the order parameter that is plotted along the abscissa is interpreted as the position of a ball in a potential landscape. Upper part: When the control parameter is below its critical value there is a stable position of the order parameter at the bottom of the valley. Middle part: Close to a critical value of the control parameter the valley becomes very flat, the order parameter is pushed around by stochastic forces. Because a restoring force is small, the fluctuations become large (critical fluctuations). Also the order parameter relaxes towards its equilibrium point only slowly (critical slowing-down). Lower part: Beyond the critical value of the control parameter two valleys occur that in many cases are symmetric. A chance event decides which of the valley will be occupied by the order parameter.

function f can be calculated to any desired degree of accuracy. From the many explicit results on the low dimensional dynamics I just quote the behavior of a single order parameter under the impact of fluctuations (cf. Figure 1). Its behavior can be particularly simply visualized. We identify the size of the order parameter with the position of a ball in a landscape that is deformed when a control parameter is changed so that an instability is reached. A typical sequence of such landscapes is shown in Figure 1 with its legend. These brief remarks on the quantitative approach may suffice here. Readers interested in details are referred to my book "Synergetics. Introduction and Advanced Topics" (2004). From the quantitative results a number of important conclusions can be drawn which have led to the following considerations.

1.3. *Phenomenological Synergetics*

As the mathematical approach shows, even a complex system can be described by its order parameters, at least in the vicinity of its instability points. There are quite a number of systems in which the dynamics of the individual parts as described by Eq. (2) is not known or not well known explicitly. However, we may assume that at least in principle such equations exist and then we may deduce that also order parameters will exist. This leads to the idea to model a complex system by means of suitably chosen order parameters. A typical and clear-cut example is provided by our modeling of the finger movement by Kelso in which he instructed subjects to move their index fingers in parallel at a certain speed. When the speed was increased and exceeded a critical value, suddenly an involuntary change to a symmetric finger movement occurred (Figure 2). Here the control parameter is obviously the prescribed finger movement frequency whereas the order parameter is described by the relative phase between the two fingers. It has been possible to model the various aspects of the phenomena of this experiment in great detail, thus reinforcing the idea that it is possible to describe even complex systems such as the behavior of a human by means of the concept of control and order parameters. Furthermore, the proof of the existence of hysteresis, critical fluctuations and critical slowing down (cf. legend to Figure 1) has led to a revolution of our understanding of movement control: this is not achieved by a motoprogram but by self-organization.

parallel   antiparallel
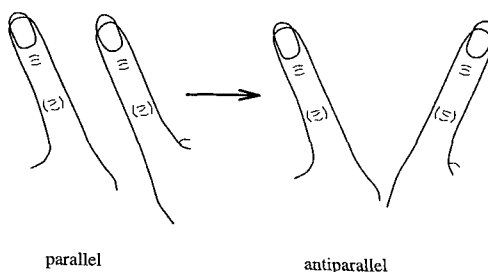
Figure 2. Spontaneous transition between two coordination patterns during index finger movement.

1.4. *Semantic Synergetics*

In a number of cases it might be difficult to model a complex system even by order parameters because quantitative measures are hardly possible. Note, however, that there is an increasing tendency to cast even qualitative features such as occurring in psychology into a quantitative frame. At any rate in the case of a lack of quantitative order parameters one may still describe a system in terms of order parameters that possess only a qualitative meaning.

1.5. *Macroscopic Synergetics*

Finally within a field I called macroscopic synergetics the goal is to derive order parameters and maybe even evolution equations by means of macroscopically given data. A description of these methods is beyond the scope of my article.

2   An Application to Psychophysics: Ambiguous Figures

The order parameter concept allows us to model the phenomenon of the perception of ambiguous figures in considerable detail [2]. An example for such a figure is shown in Figure 3. We attribute two order parameters ξ_1, ξ_2 to the percept of a young lady and an old lady, respectively, and also attention parameter λ_1, λ_2. Just to illustrate the treatment of this phenomenon in the frame of phenomenological synergetics by means of order parameters we present the corresponding order parameter equations

$$\dot{\xi}_1 = (\lambda_1 - D)\xi_1 + \xi_1^3 , \tag{6}$$

$$\dot{\xi}_2 = (\lambda_2 - D)\xi_2 + \xi_2^3 \tag{7}$$

with

$$D = \xi_1^2 + \xi_2^2 . \tag{8}$$

The crucial assumption is in accordance with an earlier proposal by Köhler to assume that attention fades away if a pattern has been recognized. Our own formulation of this saturation of attention is presented by the equations

$$\dot{\lambda}_1 = \gamma\left(1 - \lambda_1 - \xi_1^2\right), \tag{9}$$

$$\dot{\lambda}_2 = \gamma\left(1 - \lambda_2 - \xi_2^2\right). \tag{10}$$

Where γ is a suitably chosen constant. I mention that the bias which percept is realized first can be taken into account by our approach as well as the effect of chance events represented by fluctuating forces. Numerical solutions to the Equations (6)-(10) with the inclusion of bias is shown in Figure 3 where excellent agreement with experimental

results can be found by a suitable choice of the parameters. If the bias exceeds a critical value oscillations disappear and only one perception lasts forever.
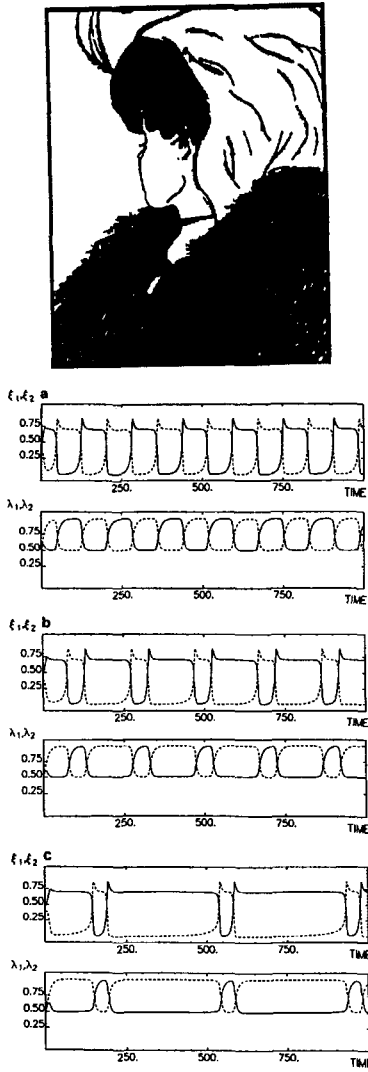
Figure 3. Upper part: Young woman or old woman? An example of an ambiguous figure showing oscillations of perception. Lower part: Oscillations of order parameters ξ_1, ξ_2 and for attention parameters λ_1, λ_2 for increasing bias (from top to bottom)

In the frame of phenomenological synergetics we have treated further phenomena such as hysteresis of perception (Figure 4,5). When one looks consecutively at a series of pictures, for e xample, i n which a certain bias is slowly changed from one value to the other one (a famous example is provided by a man's face that continuously goes over into the picture of a kneeling woman. A jump from the one perception to the other one occurs depending on the previously perceived pattern. If there is no bias, a chance event may determine which of the two possible interpretations is performed (in the terminology of synergetics symmetry breaking occurs).
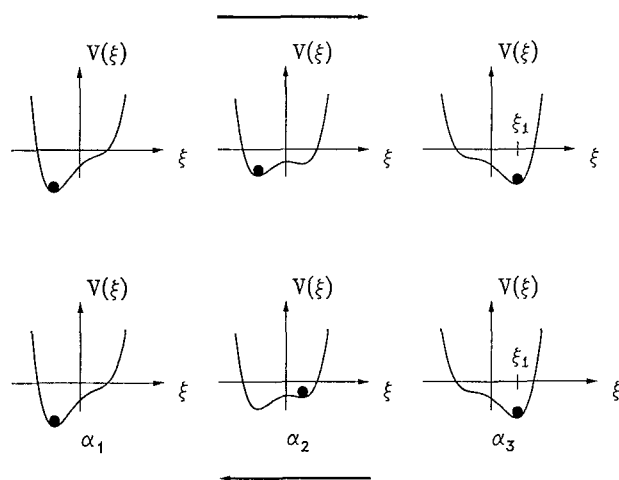
Figure 4. Description of hysteresis in terms of an order parameter the state of which is described by the position of a ball in a hilly landscape. Because of the change of the control parameter the landscape is deformed in one direction (upper part) and in the opposite direction (lower part). Note that for the same control parameter value the position of the ball that is the state of the order parameter differs because of history.
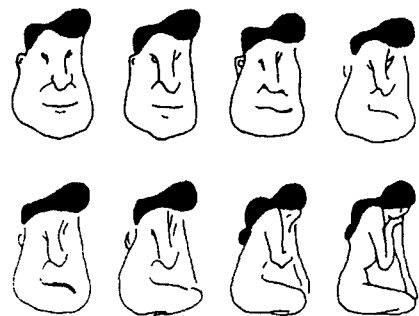
Figure 5. Hysteresis in perception. The switch from the perception of a man's face to that of a kneeling woman occurs at a different part of the total picture depending on the sequence of observations of the individual parts.

3   Decision Making as Pattern Recognition

As I have shown elsewhere decision making can be interpreted as pattern recognition. Then the following analogies can be established:

1.   Unique perception → unique decision.
2.   Two equivalent possibilities, symmetry breaking → a small chance event may determine which of the two possible choices is made.
3.   Hysteresis: our perception depends on our previous perception in spite of the fact that objectively speaking the bias has changed. → We make decisions similar to those corresponding to previous situations in spite of the fact that the situation has changed. I have called this phenomenon manager syndrome. When a manager has been successful in the past, he tends to make similar decisions in spite of the new situation.

I want to stress that all these phenomena can be modeled at the order parameter level and in a number of cases still below it (cf. the final section).

4   Health Care: Applications to Medicine and Psychology

Synergetics may shed some new light on processes dealt with in medicine and psychology as I will demonstrate below by means of a more specific case. The general idea, however, is as follows: it is based on the concept of circular causality defined by the circle between order parameters and enslaved parts that in turn generate the order parameters. This implies the stability of a system governed by one or several order parameters. It can also be interpreted as the robustness of a system against deviations of one or several subsystems from their behavior prescribed by the slaving principle. In other words, when due to a fluctuation or another cause for a certain time a subsystem does not follow the prescribed behavior it will be pulled back to its "normal" behavior. There are two important conclusions, namely

1.   When the order parameter or a set of them represent a healthy state, we, at least intuitively, understand immediately that there is a self-healing capability of the system provided enough subsystems are still in the appropriate state.

Table1. Error-Correction in reading.

Spelling Errors

Aoccdrnig to a rseearch at an Elingsh uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoatnt tihng is that frist and lsat ltteer is at the rghit pclace. The rset can be a ttoal mses and you can sitll raed it wouthit porbelm. This is bcuseae we do not raed ervey lteter by it slef but the word as a wlohe.
Preosllnay I tinhk its cmolpete Bkolcols

2. New insights for therapy become possible if the order parameters represent some illness. T he o rder p arameter c oncept then implies that not just one component but several do not act normally. In this case we put the order parameter in analogy to the concept of a syndrome. This is not meant, however, only superficially, just by listing up the individual features of the enslaved parts. In the case of a syndrome we take this analogy seriously by saying that a syndrome is a dynamical state equivalent to that of an order parameter.

In order to change an order parameter it is not sufficient to change the behavior of an individual part but to simultaneously change the behavior of a decisive number of parts so that the newly resulting order parameter can act self-consistently. At the same time if needed we also must change the appropriate control parameter values. Quite clearly specific subsystems are more important then others and, at least at the present stage, synergetics can give only qualitative hints for specific situations. While I believe that my conclusions are valid for a rather large range of phenomena in medicine, I want to concentrate somewhat more on these aspects in the field of psychotherapy.

4.1. *Can Synergetics Formalize Emotions?*

In order to achieve this goal we aim at quantitative registration according to the scheme: how often, how long, how strong is an objective or subjective "observable". Subjectively these questions can be answered by introspection or objectively by a number of quantities which are physical, chemical quantities such as blood pressure, skin resistance, hormone levels, EEG, MEG, fMRI... registrations. External observations such as facial expressions (e.g. FAX-Ekman), movement patterns, posture, kind of speech (hastily, etc.). More recently we are aiming at automatic recognition by means of the synergetic computer dealing with the interpretation of facial expressions, and language recognition. The relationship between the order parameter (s) and the individual parts follows the usual scheme where the individual parts are in our case represented by feature vectors that may contain both subjective and objective components. As mentioned above we set a syndrome in analogy to an order parameter. An example of states of order parameters is provided by joy, sadness, fear... The order parameter concept implies associative completion as exemplified by the associative memory which can also lead to error correction. An important role is played by the interaction between emotion and cognition leading to the question which one acts as control parameter, which acts as order parameter, which enslaves what. In my interpretation emotion and cognition condition each other which is related but not identical with Ciompi's concept of affect logics which is so-to-speak an order parameter that is related to affects on the one hand and logics on the other hand, whereby in the parlance of synergetics affects and logic play the role of enslaved subsystems.

Let us discuss a few consequences for psychotherapy. A simple but rather instructive picture of the desired effect of intervention is as follows: we refer to Figure 4. On the left hand side the client is characterized by his/her order parameter in the left valley which is

an undesired situation, whereas the desired situation is indicated by the upper valley. How is it possible to bring the client from the left to the right position? A strategy offered by synergetics is as follows: by changing appropriate control parameters the landscape is deformed from the left hand side to the right hand side and then by further interventions the state of the person is brought from the left hand state to the right hand state. Note, however, that in the case of the right hand side in a number of cases external interventions are no more needed. The transition occurs by means of self-organization. In general the situation is more complicated because the state of the system must be described by means of a multi-dimensional picture, say, a two-dimensional landscape. Now destabilization must occur only in specific directions whereas the other directions must be kept stable. In addition in a number of cases, even in the newly destabilized state new more or less equivalent stable states (valleys) become accessible, some of which are still undesired. In such a case the symmetry-breaking must be done by external interventions. Let me be more specific. In order to destabilize a landscape and transform it to a new one in a number of cases antipsychotic drugs may be administered and may be more efficient than verbal interventions because the latter act by learning via Hebbian synapses which is a slow process. However, once antipsychotic drugs are administered, still the possibility of symmetric states arises, some of which are unwanted. In such case the newly evolving valleys must be equipped with meaning which can occur only via the psychological interventions.

5   Synergetics in Management

Synergetics as outlined in the first section as well as numerous examples treated at various levels provide us with a number of insights into management. Synergetics can be considered as a theory of indirect control rather than influencing the individual parts of a system directly and in a specific manner. Rather, the parts are controlled only in a global way. Nevertheless by means of self-organization highly ordered structures with interesting properties evolve. The example of the light source laser may serve as a nice paradigm for the effects of self-organization. The essential components of a laser are light emitting atoms or molecules. After being excited by some energy influx, e.g. an electric current, they emit light waves that in a usual lamp are entirely uncorrelated. When the energy input, that serves as control parameter, exceeds a critical value, a highly ordered light wave is generated via self-organization. Beyond this critical control parameter value, the efficiency of the system is changed abruptly (cf. Figures 6 and 7). Already this effect may be an incentive to study in more detail the mechanisms by which this efficiency is increased. First of all the appropriate atoms and molecules must be selected having specific properties. In other words, the laser self-organization is based on a number of specific properties of the components. They must emit their light waves strongly enough and be able to interact with the other atoms sufficiently strongly. Furthermore they must

be in an adequate surrounding, e.g. embedded in-between two parallel mirrors that reflect the generated light sufficiently often.
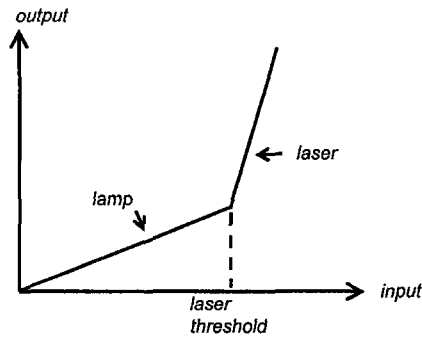
Figure 6. Laser output versus energy input. Note the sharp transition at the laser threshold.
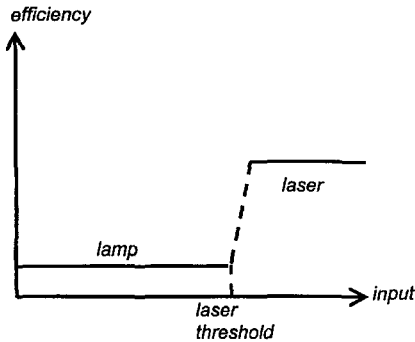
Figure 7. The differential efficiency is defined by the size of the slope of the curve of Figure 6. Note the large jump.

Even without going into any further mathematics one can draw a number of qualitative conclusions with respect to management. First of all the choice of capable coworkers is essential. They must be willing to cooperate in a productive fashion and they must be brought in a stimulating environment. How can one attract in practice the appropriate coworkers? This requires a good standard for the life of their families such as schools, shopping centers and further amenities. Furthermore we must enhance the cooperation between coworkers within the company as well as outside of it. An important aspect of laser physics is the ability of the atoms and molecules to emit light spontaneously. This means that coworkers must be able to develop new ideas spontaneously. Correspondingly to the laser case there must be a transition period during which by brain storming new ideas are produced, checked against each other, realized by means of pilot projects and eventually selected (Figure 8). As the mathematical treatment of the generation of laser light shows, during the transition period there may be considerable fluctuations in the output (Figure 9). It takes a certain time until the coherent laser wave with its high

efficiency is established. This has quite evidently practical consequences also in management. During the transition period we must allow that errors are made and we must allow a for minimum time during which the new efficient system is established. I do not think that this is a superficial analogy but it points at an important issue, namely that in management we must observe fundamental laws of system dynamics. Careful planning surely can attempt to minimize errors or failures but in view of the complexity of modern systems there are limits.
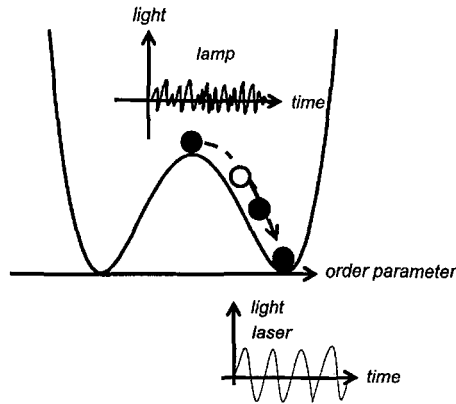
Figure 8. In the hilly landscape (lower part of Figure 1) the order parameter must roll down from the top to the bottom, whereby the disordered light of the lamp goes over into the fully ordered light of the laser with its high efficiency. Note that this transitions takes some specific time.
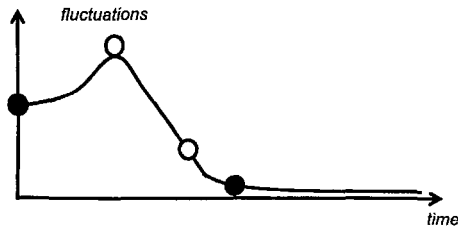
Figure 9. Development of the fluctuations of laser light during the transition of the order parameter shown in Figure 8.

Let me summarize my brief excursion into management. The basic concept of synergetics lies in the indirect steering of systems, i.e. the task of managers should be to generate general conditions under which self-organization becomes possible. This includes, of course, the appropriate provision of material and also immaterial resources, the establishment of corporate identity as the uppermost order parameter so that coworkers

can identify themselves with the goals of the company, sufficient interaction between coworkers, flat hierarchies in order to avoid the information bottleneck that has been well-known especially in Eastern countries before the breakdown of the Soviet economic system, etc.

6   Outlook

In my article I have tried to indicate some developments of synergetics with respect to what is usually called "the soft sciences". While synergetics started from physics and has found a profound mathematical basis, there is the tendency, even in the soft sciences, towards quantification and mathematical modeling. I think that synergetics may be of help along this way where starting from the level of semantic synergetics we may proceed to modeling by means of order parameters and eventually to a microscopic approach.

A few hints may be useful. In psychology and psychiatry attempts at quantification have been successfully made in the past. More recently psychological and psychiatric processes have been analyzed in terms of time series. Analyses that are well-known to the science community dealing with fractals have been performed and predict the change of a mental state of a person. By means of quantitative analysis also the effects of the interaction between therapists and clients could be elucidated, especially by the work of Schiepek and coworkers (For references cf. [3]). Furthermore with respect to cognition, synergetics can provide us with new insights in particular by use of the order parameter concept in recognition processes of patterns and decision making. I think here the mathematization has progressed most, especially based on simple models of neurons as well as by use of computers and there is a vast literature which to referee here is beyond the scope of my paper. However, I want to indicate that concepts of synergetics have proven very useful in order to establish realistic models of neural networks and especially to cut through the very complicated equations. I just want to spend only a few words on these recent developments. The starting point is a rather realistic model of a neuron composed of its dendritic tree, its soma and its axon (Figure 10). The dendritic tree receives input signals from other neurons via synapses. These input signals are converted into dendritic currents that in turn within the soma are converted into voltage pulses that travel along the axon which then splits and makes contact to the dendrites of other neurons via synapses. This leads to the formulation of a pulse-coupled neural network in which for the first time the coupling between different neurons has been taken into account in a realistic fashion so that it has become possible to treat not only pulse-synchronization but also pattern recognition [4]. The key to an appropriate pattern recognition algorithm consists in a suitable transformation of the original microscopic equations into order parameter equations that directly refer to the patterns to be recognized. At the same time this time dependent mechanism of the establishment of order parameters on the one hand and the fading away of attention parameters on the other hand leads to a detailed theory of a finite window of perception that is perceived as

present. In this way the experimental findings of Pöppel and others on a finite time window that is interpreted as present can be given an explanation.
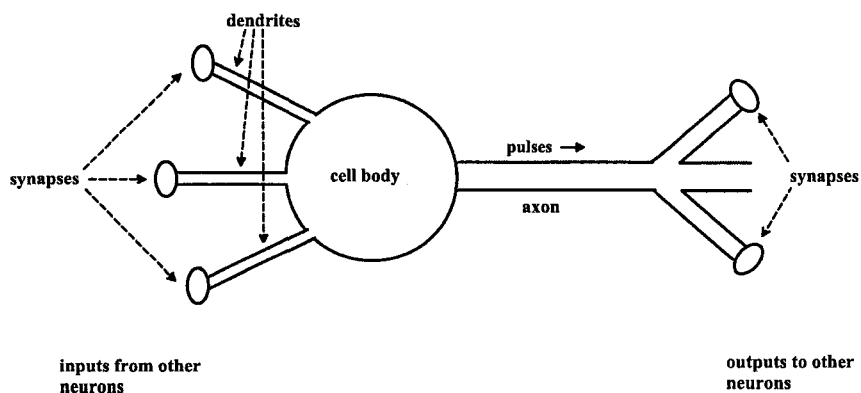
Figure 10. Model of a neuron with its dendrites and synapse (cf. text). By an appropriate choice of the synaptic strengths the corresponding neural network can model both pattern recognition as well as synchronization of the firings of neurons.

In summary I may say that the methodology of synergetics seems to be capable of providing new insights and theoretical approaches to the life sciences.

References

1. H. Haken, *Synergetics and Advanced Topics*. Springer, Berlin (2004).
2. H. Haken, *Synergetic Computers and Cognition*. Springer, Berlin (2004).
3. H. Haken and G. Schiepek *Synergetik in der Psychologie*. Hogrefe, Göttingen (to appear)
4. H. Haken, *Journal of Psychophysiology*. (to appear).

This page is intentionally left blank

COMPLEXITY, FRACTALS, NATURE AND INDUSTRIAL DESIGN: SOME CONNECTIONS

NICOLETTA SALA

Accademia di Architettura
Università della Svizzera italiana,
Mendrisio, 6850, Switzerland

Industrial Design is the field of developing physical solutions which might include products, vehicles, machinery, and even environments. For many years the industrial designers found inspiration by the symmetry and by the Euclidean shapes (e.g., polygons and polyhedra). The evolution of the technologies (from the hand-made to the Computerized Numeric Control) and the evolution of the materials (from glass to the polymer composites) have permitted to the designers to conceive shapes which pass the limits imposed by the Euclidean geometry. Modern design studies apply complex shapes based on the fractal geometry to create new kind of objects, realized in innovative materials. These objects sometimes mimic the complex shapes present in the nature. The aim of this paper is to describe an approach which introduces the connections between the Nature and some industrial design objects. Their complex shapes will be analysed using a fractal point of view.

1 Introduction

In the 20th Century, there was an important transformation of our society from that of hand-made consumer goods realized and made by skilled craftsmen to one of mass-production with new materials and technologies. This transformation influenced the coming of the Industrial Design, which is a field of design that emerged in the late 1920s and early 1930s. The industrial designers can find inspiration by mathematics components. The Euclidean geometry, the symmetry, the golden ratio, the proportions, the curves (e.g., Bézier curves and B-splines) and the surfaces help the designers to create a canon of beauty [1, 2, 3, 4, 5]. We can analyse the industrial design objects using a different point of view, for example we can observe some complex or fractal components. In our analysis the complexity is not conceived as an area of chaos, but as an attempt to overcome the shapes imposed by the Euclidean geometry. The complexity is also connected to the fractal geometry. Fractal geometry describes the irregular shapes and it can occur in many different places in both Mathematics and elsewhere in Nature. Fractal objects are irregular in shape, and they are generally self-similar and independent of scale. There are several mathematical structures that are fractals; for instance the Julia set [6], the Koch snowflake [7], the Mandelbrot set [6] and the Lorenz attractor [8]. Complexity and fractal geometry can inspire a new aesthetic sense [2, 9]. In the 1930, the American mathematician George Birkhoff (1884-1944) proposed a measure of beauty defined as:

$$M = \frac{O}{C} \tag{1}$$

whereby M stands for "aesthetic measure" (or beauty), O for order and C for complexity. This measure could suggest the idea that beauty has something to do with order and complexity, but the Birkhoff's relationship is not used by the designers.

Our studies followed these two questions:

1. "Why the complex and fractal components are present in the industrial design objects?"

2. "Can the complexity and the fractals help to create a new kind of industrial design objects?"

To answer to the first question, it is important to consider that complex and fractal components are present in the industrial design to reproduce the Nature, which is visually complex and fractal-like [10, 11]. In the industrial design this reproduction has influenced the development of materials, for example the introduction of the float glass, and new manufacturing techniques (for example the use of special laser).

To answer to the second question our analysis has been organized in two parts:

• To find the complex components connected to the natural shapes which have inspired some industrial design objects (e.g., turbulent motions of the water);

• To observe the fractal components in the industrial design objects, in particular the self-similarity (for example in the Hadid's objects).

2 Complex Components In The Industrial Design Objects

Industrial design find inspiration by the complexity of the Nature, and a new kind of industrial design objects were born observing the complex components connected to the natural shapes and natural phenomena. The chaotic movement of the water has inspired the internationally renowned Welsh born designer, Ross Lovegrove. He is famed for his futuristic, organic minimalist designs and is one of this century's leading international designers. Lovegrove's passion for natural shapes, for example the water motion, influenced the creation of the visionary bottle for the spring water TY NANT (2001), shown figure 1. It has been produced using a plastic resin of the polyester family (PET: Polyethylene terephthalate).

Figure 1. TY NANT bottle (2001) recalls the water motions.

The German lighting designer Ingo Maurer, that raised lighting design and lamps to a high art form, designed the lamp *Paraguadì* (1997), shown in figure 2, for Guadi's *Casa Botines* (1891-1892). Probably, he found inspiration for this lamp observing the turbulent motions of the smoke.

Figure 2. The lamp *Paraguadì* (1997) by Maurer could be inspired to the turbulent motions of the smoke.

The Italian architect and designer Alessandro Mendini cultivates a particular interest in neo-modern and contemporary design. His armchair *Proust* (1978) has a complex cover, shown in figure 3a, that recalls the Pollock's painting *Blue Poles: Number 11* (1952) which is manifestly fractal [9].

a)   b)

Figure 3. *Proust* (1978), realized by Mendrini for Studio Alchimia (Italy) a), shows a complex cover that recalls the Pollock's painting *Blue Poles: Number 11* (1952) which is fractal b).

In 2001, Alberto Alessi, the third-generation owner of his family's Italian manufacturing company, commissioned twenty-two internationally acclaimed architects to design the "Tea and Coffee Towers" project, a new landscape of objects designed by contemporary architects. This project combines the expression of architecture with the language of design [12]. Among those architects there are celebrity names such as Zaha

Hadid, Toyo Ito and Massimiliano Fuksas. All these designers use computerized techniques, fractal and non-linear geometry, and they have not a stereometric vision but a media vision of the space. They emphasize a communicative and existential fruition of the objects, obtained by original and visual components. For example, Iraqi-born British designer Zaha Hadid conceived a coffee-tea pot in 925/1000 silver for Alessi's "Tea and Coffee Towers" project, shown in figure 8. It recalls futuristic and fractal shapes. Hadid considers her industrial design objects as an "architecture in small scale", and she applies to design them the same fractal rules used in her architectural projects [9]. Observing figure 4, we can note the complexity of the shapes as a 3D puzzles, the single parts form the unity. This object exploits the combination and the contrast between the vertical and the horizontal parts [12, 19].
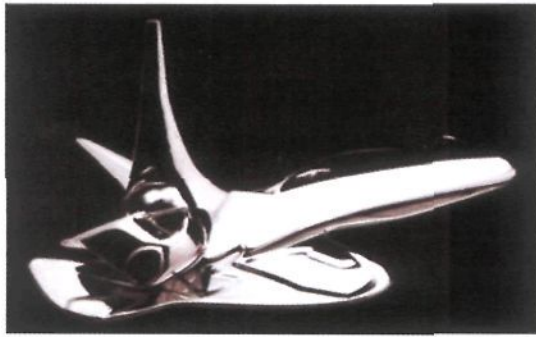
Figure 4. Tea and coffee service (2003), realized by Hadid for Alessi exploits the combination and the contrast between the vertical and the horizontal parts.

The Brazilian brothers Humberto and Fernando Campana design objects with complex geometry. They find your inspiration observing the shapes, particular complex, present in the Nature (e.g., trees, anemones and corals). For example, the citrus basket and the centrepiece bowl *Blow up* (2004), realized for Alessi in 18/10 stainless steel mirror polished, shown in figure 5a and 5b, and the chairs in the figures 6 and 7 were inspired by the natural shapes.

a)   b)
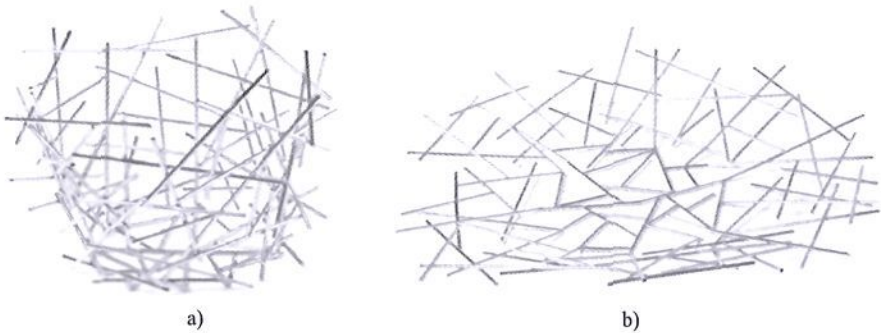
Figure 5. Citrus basket *Blow up* (2004) a) and centrepiece bowl *Blow up* (2004) designed by Humberto and Fernando Campana for Alessi.

<div align="center">a)   b)</div>
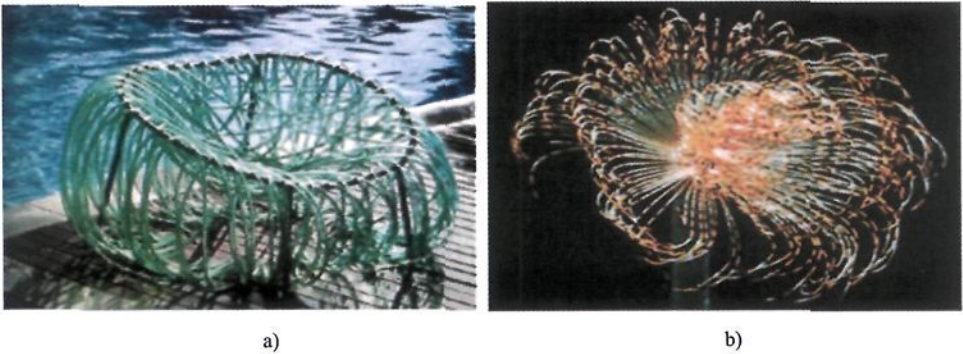
Figure 6. *Anemone chair* (2000) designed by Humberto and Fernando Campana for Edra (Italy) a), recalls the complex shapes present in the sea (Spirographis Spallanzani) b).

<div align="center">a)   b)</div>
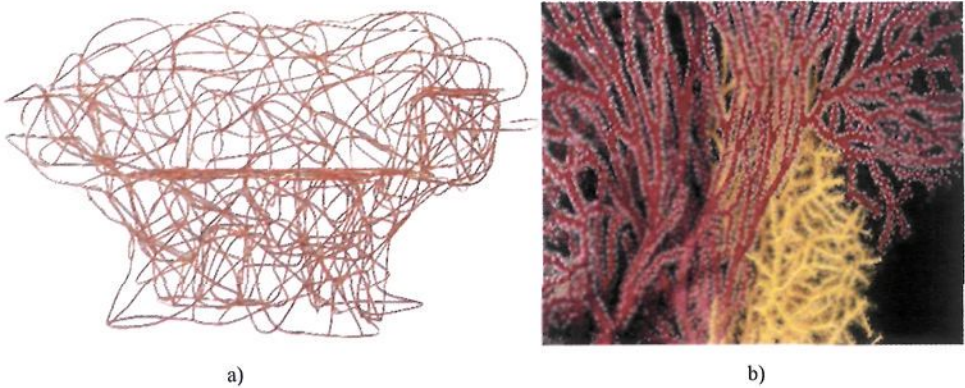
Figure 7. *Coral Chair* (2003-2004) by Humberto and Fernando Campana for Edra a), mimics the branching structures of the corals b) which can be studied using a fractal approach [13].

3 Fractal Components In The Industrial Design Objects

Fractal geometry is a mathematical theory that radically departs from traditional Euclidean geometry. It describes objects that are scale symmetric, or self-similar. This means that when such objects are magnified, their parts are seen as an exact resemblance to the whole, the property continues with the parts of the parts and so on to infinity. These shapes are called "fractals", and they must maintain a rough, jagged quality at every scale at which an object can be examined. The nature and the characteristics of fractals are reflected in the word itself, coined by the Polish-born French mathematician Benoit B. Mandelbrot from the Latin verb *frangere*, "to break", and from the related adjective *fractus* that means "fragmented and irregular" [14, 15, 16, 17]. The acceptance of the word "fractal" was dated in 1975. Objects that are now called "fractals" were discovered and explored long before the word was coined (e.g., the Koch snowflake). A fractal object is self – similar if it has undergone a transformation whereby the dimensions of the structure were all modified by the same scaling factor. The new shape

may be smaller, larger, rotated, and/or translated, but its shape remains similar. "Similar" means that the relative proportions of the shapes' sides and internal angles remain the same. The self-similarity is defined by Mandelbrot as follow: "When each piece of a shape is geometrically similar to the whole, both the shape and the cascade that generate it are called self-similar" [10, p. 34]. Mandelbrot used the term "self-similar" for the first time in 1964, in an internal report at IBM, where he was doing research, and in the title of a 1965 paper. It is possible to demonstrate that the fractal shapes and the self-similarity were known to the artists in different cultures [18].

Industrial designers use the fractal geometry as a tool which permits to overcome the limit imposed by the Euclidean geometry. They can design new complex shapes to create a new kind of industrial design objects inspired by the Nature, based on the repetition in different scales, self-similarity, and on the "broken" symmetry [19]. For example, the American artist Danny Lame realized *Shell* (1988) for FIAM (Italy), shown in figure 8a. It is a coffee table with hand-sculptured float glass top of 15 millimetres thick. The base is made of five 12 millimetres thick hand-sculptured float glass elements, four of which curved. All the elements are kept together by stainless steel hardware. Lame's coffee table is a clear example of a design object that mimics the fractal components present in the nature, in particular: the "broken" symmetry and the self-similarity present in the shells (figure 8b).

Other industrial design object which has a fractal-like shape is the coffee table *Papilio* (1985) realized by Mendini for Zanotta (Italy). The table has two or three shelves in 8 millimetres, shaped and hand-cut tempered glass. Legs in brushed, nickel plated steel, screwed onto stainless steel discs fixed to the glass. Figure 9a shows the three shelves version of this coffee table, and figure 9b illustrates the vision of the top that emphasizes the irregularity [19].

The repetition in different scales is also present in the Alessi's "Tea and Coffee Towers" project. The architect and designer Tom Kovac, Australian of Slovene descent, realized for Alessi a tea and coffee service in double wall 925/1000 silver, shown in figure 10. Kovac is the Creative Design Director of Curvedigital a joint project between RMIT University and Melbourne Museum. He has explored the potential of new processes and developments in non standard manufacturing. The set of the objects that composes this tea and coffee service is a kind of code which activates a geometric progression, shown in figure 11. This progression opens new hypothesis in the design's methodologies. The parts that form the set derive by correlation criteria that use the inference by the conceptual to the operative [12, 19]. Observing figure 11 we note that the same shape is repeated in four different scales.

The self-similarity is also present in the cover of the chairs, an example, realized by the designer Paul Smith, is shown in figure 12a. The figure 12b illustrates a particular of a Mandelbrot set of the level $\alpha(c)$ for $c \in M$, based on the Alfred Barten's experiment [7]. The analogy between the figures 12a and 12b reveals that the designers' creativity can find inspiration by the self-similar shapes connected to the fractal sets.

a)   b)

Figure 8. The coffee table *Shell* (1988) a) mimics the self-similar shapes present in the nature b).

a)   b)
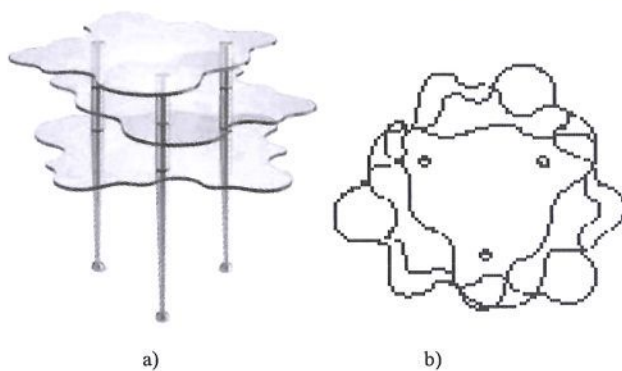
Figure 9. Coffee table *Papilio* (1985), by Mendini for Zanotta emphasizes the self-similarity a) the vision of the top shows the fractal components b).
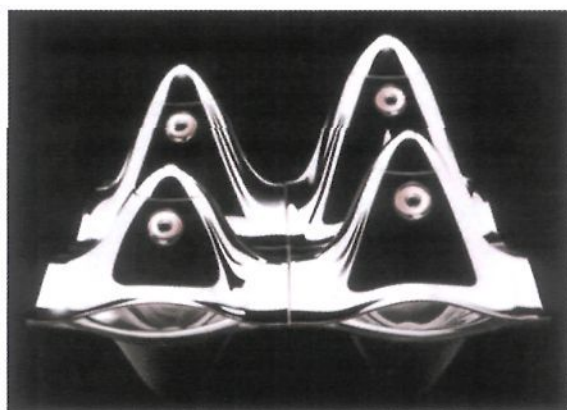
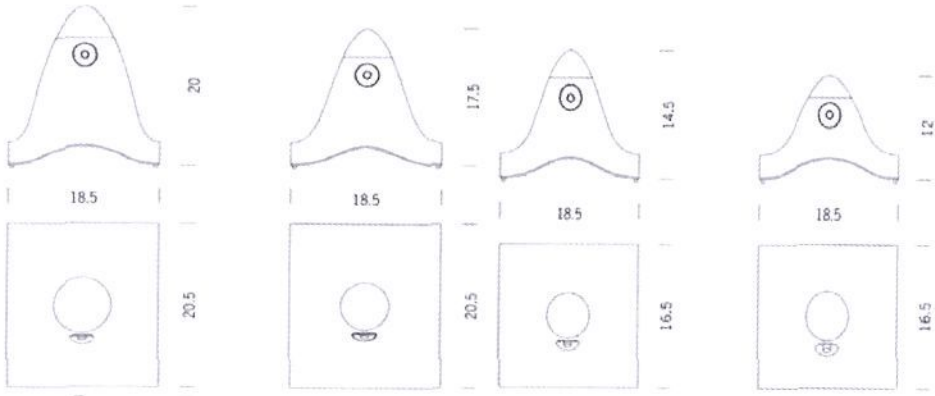Figure 10. Tea and coffee service (2003) realized by Kovac for Alessi.

Figure 11. Tea and coffee service (2003) realized by Kovac for Alessi shows the same shape repeated in four different scales.

a)   b)
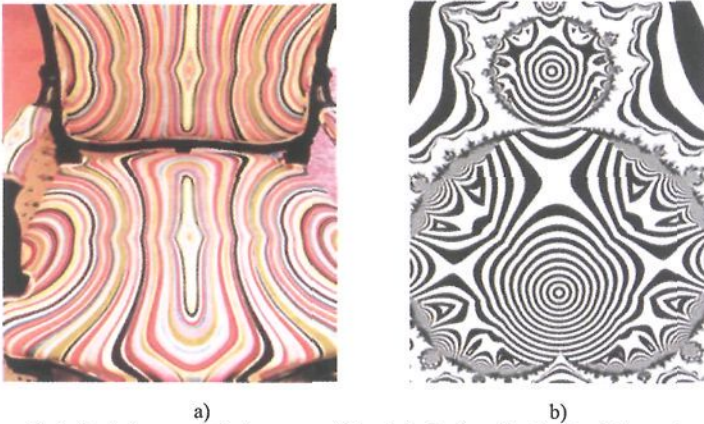
Figure 12. The self-similarity is present in the cover of the chair designed by Smith a) the analogy with the fractal set b) is interesting.

Natural and fractal patterns are used in the realisation of the industrial design objects. Frank O. Gehry created his playful *Fish and Snake Lamps* (1983-86) observing fishes' and snakes' skins that have self-similar pattern. Gehry designed the lamps that would emphasize the translucency of the laminate's integral coloration. The serendipitous discovery of splintered patterns came after several failed attempts at a design in which one lamp was accidentally broken. Gehry translated the scale-like pattern of the jagged pieces into fish and snake forms. Another example of a lamp inspired by self-similarity presented in the nature is shown in figure 13. It is *Styrene* (2003) conceived by the British lighting designer Paul Cocksedge that creates visually spectacular and technically ingenious lights that celebrate the magical and transformative qualities of illumination. *Styrene* has an interesting analogy with the pangolin's skin (in figure 13c).

a)   b)

c)
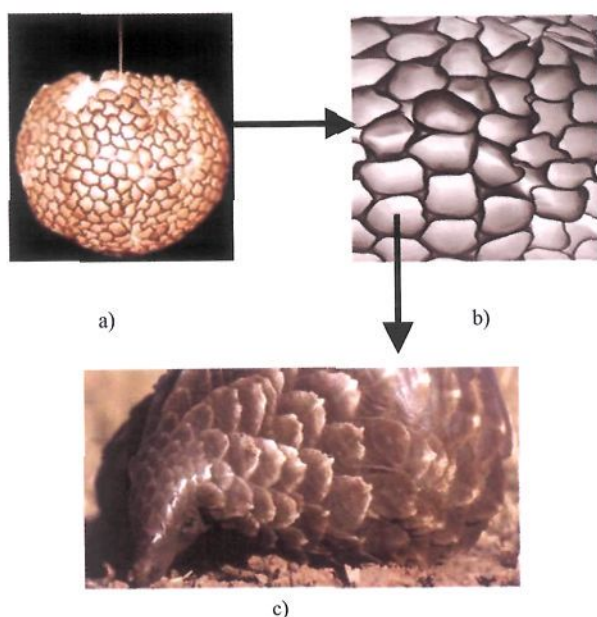
Figure 13. *Styrene* (2003) by Cocksedge a) evidences a kind of self-similarity b) which is also present in the Pangolin's skin c).

4   Conclusions

Complexity, fractals, nature, and industrial design have some interesting connections. The complexity paradigm is developing three different traditions: in the design fields (architecture, engineering, and industrial design), in the arts, and in the science [9, 18, 20, 21, 22]. This work describes only an approach where the complexity and the fractal geometry help the designers to create new kinds of industrial design objects, based on iterative processes. The fractal geometry is a potent language of complex visual form, in fact it reduces much of the staggering complexity present in nature to some very simple mathematical rules. The examples described in this work introduce the concept of non-linear design. In the Alessi's "Tea and Coffee Towers" project, here described, there is an isomorphism between the complexity in architecture (complexity in the large scale) and the complexity in the industrial design (complexity in the small scale). All designers involved in this project worked on a dizzy change of scale and on some very different functions [12, 19].

In this century, the joint evolutions of materials, manufacturing processes and design can create great products that have complex shapes [21, 22, 23, 24, 25, 26]. The complex components closely connected to the fractal geometry will define new CAD (Computer Aided Design) systems, and they will influence the industrial design in the next future.

References

1. I. Hargittai, and M. Hargittai, *Symmetry a unifying concept*, Shelter Pubblications, Inc: Bolinas, California (1996).
2. U. Eco, *Storia della bellezza*, Bompiani, Torino (2002).
3. B. Munari, *La scoperta del triangolo*, Zanichelli, Bologna (1976).
4. B. Munari, *La scoperta del quadrato*, Zanichelli, Bologna (1987).
5. K. Elam, *Geometry of design*, Princeton Architectural Press, New York (2001).
6. H. von Koch, Sur une courbe continue sans tangente, obtenue par une construction géometrique élémentaire. Arkiv för Matematik 1, pp. 681-70 (1904).
7. H-O. Peitgen and P.H. Richter, *The Beauty of Fractals: Images of Complex Dynamical Systems*, Springer-Verlag, Berlin (1986).
8. J. Crilly, R.A. Earnshaw and H. Jones, *Fractals and Chaos*, Springer-Verlag, New York, pp. 43-69 (1991).
9. N. Sala and G. Cappellato, *Architetture della complessità*, Franco Angeli, Milano (2004).
10. B.B. Mandelbrot, The fractal geometry of nature, W.H. Freeman, San Francisco (1982).
11. D.R. Fowler, H. Meinhardt, and P. Prusinkiewicz, Modelling seashells, *Proceedings of SIGGRAPH '92: Computer Graphics*, 26 (2), pp. 379-387 (1992).
12. A. Mendini, *Tea & coffee towers*, Electa, Milano (2003).
13. E. Basillais, Coral surfaces and fractal dimensions: a new method, *CR Acad. Sc. Paris, Life Sci.*, 320, pp. 653-657 (1997).
14. B.B. Mandelbrot, *Les objects fractals. Forme, Hasard et Dimension*, Flammarion, Paris (1975).
15. H.O. Peitgen, H. Hurgens and D. Saupe, *Chaos and Fractals: New Frontiers of Science*, Springer-Verlag, New York (1992).
16. R. Fivaz, *L'ordre et la volupté*, Press Polytechniques Romandes, Lausanne (1988).
17. J. Briggs, *Fractals the patterns of chaos*, Thames and Hudson, London (1992).
18. N. Sala, Fractal Geometry In The Arts: An Overview Across The Different Cultures. Novak M.M. (Eds.) *THINKING IN PATTERNS Fractals and Related Phenomena in Nature*, World Scientific, Singapore, pp. 177-188 (2004).
19. N. Sala and M. Sala *Geometrie del design. Forme e materiali per il progetto*, Franco Angeli, Milano (2005).
20. J.A. Kaardorp, Modelling Growth Forms of Sponges with Fractal Techniques, A.J. Crilly, R.A. Earnshaw, and H. Jones, *Fractals and Chaos*, Springer-Verlag, New York, pp. 71-88 (1991).
21. P. Ball, *The Self Tapestry. Pattern formation in nature*, Oxford University Press. Oxford (1999).
22. A. Bejan, *Shape and Structure from Engineering to Nature*, Cambridge University Press, Cambridge (2000).
23. A. Bangert and A. Karl, *80's Style: Designs of the Decade*, Abbeville Press, New York, (1990).
24. S. Bayley (ed.), *The Conran Directory of Design*, Villard Books, New York (1985).
25. J.de Noblet (ed.), *Industrial Design: Reflections of a Century* Flammarion/APCI, Paris (1993).
26. M. Ashby and K. Johnson, *Material and Design*, Butterworth Heinemann, UK (2002).

MODELLING PATTERN FORMATION UPON LASER-INDUCED ETCHING

A. MORA, M. HAASE

Lehrstuhl für Höchstleistungsrechnen, University of Stuttgart,
Nobelstr.19, 70569 Stuttgart, Germany
E-mail: mh@ica.uni-stuttgart.de

T. RABBOW, P.-J. PLATH

Institut für Angewandte und Physikalische Chemie, AG Chemische Synergetik,
University of Bremen, Bibliothekstr. NW2, 28334 Bremen, Germany
E-mail: plath@uni-bremen.de

H. PARISCH

Institut für Statik und Dynamik der Luft- und Raumfahrtkonstruktionen,
Pfaffenwaldring 27, 70569 Stuttgart, Germany
E-mail: parisch@isd.uni-stuttgart.de

In this paper we present a discrete stochastic model for laser-induced jet-chemical etching in one and two spatial dimensions for the microstructuring of metals. Experimental observations show that depending upon the choice of system parameters either surface roughening occurs or the formation of periodic ripple structures. The results of kinetic Monte-Carlo simulations for the discrete model are compared to experimental observations. The continuum limit taken for the discrete model leads to a modified Kuramoto-Sivashinsky (KS) equation describing the evolution of the etching front. This type of equation is also characteristic for other erosion processes like water-jet cutting and can be considered as an order parameter equation. The spontaneous ripple formation can then be explained by the development of a front instability. Finite element simulations of the modified KS equation support the theoretical findings.

Keywords: discrete stochastic model, laser-induced etching, Kuramoto-Sivashinsky equation, finite elements

1. Introduction

Recently, considerable effort has been devoted to improve laser-induced processes for microstructuring of metals[1,2,3,4]. Laser-induced jet-chemical etching processes are of growing interest for microstructuring, because the low-power laser beam in combination with an etchant jet allows a highly localized material removal without imparting thermal or mechanical stresses on the workpiece. Metallic microparts of high surface quality are used for example as biomedical tools or as moulds for integrated optical microsystems. The laser-induced jet-chemical etching technique

has been successfully applied for the fabrication of superelastic microgrippers cut out of temperature-sensitive shape memory alloys[2]. However, under certain process conditions different surface morphologies are observed like roughening or the formation of unwanted ripples. These ripple structures may either be triggered by some external periodic forcing or they may be generated intrinsically due to the formation of an instability[5]. In order to avoid such parameter regimes and to optimize the process, a profound understanding of the basic mechanisms underlying the dynamics of the interacting thermochemical and hydrodynamical processes which lead to spontaneous pattern formation is necessary.

Periodic patterns can be observed ubiquitously in nature and technical processes on a large range of scales. In nature, well known examples are the formation of cloud streets or the aeolian ripples in sandy deserts[15]. In technical systems, ripples occur during ion sputtering of amorphous surfaces[9], in laser beam cutting[10] and water-jet cutting[14,5] or in laser ablation of transparent dielectrics[11]. Common to all these systems is, that they are nonlinear open systems operating far from thermodynamical equilibrium in which at least two counteracting mechanisms exist, a destabilizing dynamics which is balanced by a stabilizing one. Since in laser-induced jet-chemical etching the interaction of thermodynamical, chemical and hydrodynamical processes is highly complex and not known in all details, we introduce here a phenomenological discrete stochastic model, which is based on a competition of erosion and diffusion processes. Despite its simplicity the model captures essential features of the process like surface roughening and ripple formation.

The article is organized as follows. In section 2 we give a short description of the principle scheme of the new laser-induced etching technique and present typical surface morphologies which are observed experimentally. The one-dimensional discrete solid-on-solid model is introduced in section 3 together with results from kinetic Monte Carlo simulations. In section 4 we briefly review the main steps necessary to derive a continuous Langevin equation for the surface evolution of etched kerfs associated to the discrete model. A linear stability analysis throws light on the spontaneous formation of ripples. An extension of the discrete model to two spatial dimensions is presented in section 5 followed by finite element simulations of the corresponding deterministic evolution equation which has the form of a modified Kuramoto-Sivashinsky equation. The results of the simulations are in good qualitative accordance with experimental observations. Finally, section 6 contains our conclusions and proposals for future work.

2. The laser-induced jet-chemical etching technique

Recently, a refined laser-induced wet chemical etching technique has been developed in which a relatively low-power laser beam is directed perpendicular to a workpiece together with a coaxial jet of fresh etchant liquid[2,4]. The metallic sample which has to be microstructured is submerged in the acid and mounted horizontally onto computer-controlled xy-stages allowing a relative movement of the sample with re-

spect to the laser beam. The absorbed light of the focussed laser beam thermally activates the surface of the workpiece, removes the passive layer and initiates etching. The main advantages of this method are the high localization of material dissolution needed for microstucturing, the enhancement of chemical reactions due to the supply of fresh etchant and removal of reaction products by the etchant jet and the low thermal loading of the workpiece. Details of the experimental setup are described elsewhere[4,7].

In order to investigate the dynamics of the process, a wide variety of kerfs has been etched for varying process parameters[4] leading to rough surface morphologies or ripple structures[6]. Essential process parameters are found to be the laser power, scan velocity, concentration of etchant and jet velocity. The ripples may either be triggered by some external periodic forcing like the etchant pump or a control mechanism of the xy-stages[4]. Alternatively, they can be generated intrinsically due to a spontaneous formation of a front instability[4,7]. In fig.1 a series of kerfs obtained for increasing scan velocities u is shown. For $u = 2\mu m/s$ a narrow uniformly etched kerf is obtained, which, however, turns suddenly for $u = 3\mu m/s$ into periodic ripple structures. For increasing scan velocities u the wavelength of the ripples decreases until, for $u = 10\mu m/s$, a broad kerf without visible periodicity is obtained.
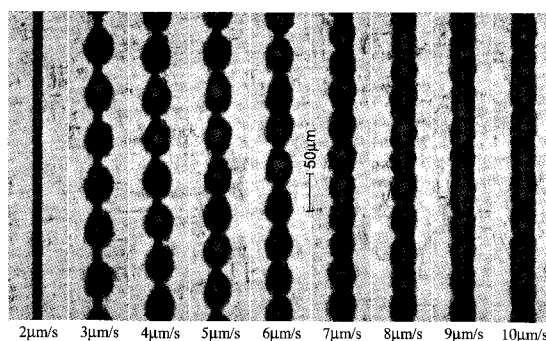
Figure 1. Microscopic image of a series of kerfs (top view) etched with increasing scan velocity u (etchant 5M H_3PO_4, argon ion cw-laser with intensity $450mW$, jet velocity $v_j = 1.59m/s$).

3. The discrete stochastic solid-on-solid model

Since we have only limited knowledge of the details of the dynamical processes involved in laser-induced jet-chemical etching, discrete models are very instructive since they represent a link between theory and experiments and allow to distinguish between essential ingredients of the interacting processes and secondary effects[8]. The discrete model which is presented in this section is a modification and adaptation of a discrete solid-on-solid model which was originally developed by Cuerno and coworkers to simulate morphologies of surfaces uniformly eroded by ion sputtering[12]. In this model periodic ripples occur at early stages at the onset of instability. Later, nonlinear effects lead to a degradation of ripple structures and a

cross-over to a rough surface morphology.

In contrast to uniform sputtering for laser-induced etching the laser beam is moved relative to the workpiece with constant velocity u thus activating chemical reactions on the surface only locally and during a characteristic time interval

$$\Delta T = \frac{d}{u} \tag{1}$$

where d denotes the diameter of the thermally activated spot. Nevertheless, the basic mechanisms are similar to ion-sputtering. There are two counteracting processes, the removal of material due to etching which destabilizes the surface, and a diffusion which has a stabilizing effect. The basic idea for ripple formation is now the following. The scan velocity u has to be chosen small enough such that ripples have time enough to develop. On the other hand, the laser should leave the active zone early enough as not to degrade the ripple structure again.

In brief, the model assumptions are the following. The laser beam thermally activates the surface, where the absorption rate depends upon the incident angle. The passive layer is then locally removed and wet-chemical etching starts. The etchant jet supplies fresh etchant and enhances the transport of heat and reaction products. In addition, multiple reflections within troughs lead to higher temperatures in valleys than on hills. These reasons cause troughs (as long as they are not too deep) to be eroded faster than mounds, i.e. the etching rate depends upon the surface curvature. In this first model we neglect the formation of a diffusive (Nernst) layer as well as an additional heat source due to exothermic reactions.
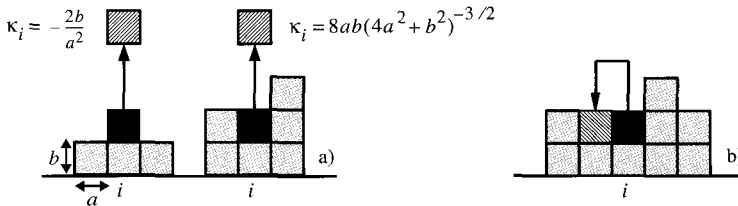
Figure 2. Discrete model for laser-induced jet-chemical etching. Schematic illustration of the curvature dependent rule a) for etching and b) for thermally activated diffusion.

In the $1+1$-dimensional case the model is defined on a lattice of rectangular cells with side lengths a and b, see fig. 2. In lateral direction we choose L cells. The interface is assumed to be periodic with heights $h_i(t)$, $i = 1, \ldots, L$. Chemical etching and thermally activated diffusion locally take place in an active zone for which we assume a Gaussian profile with standard deviation σ

$$G(x,t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-x_0)^2/2\sigma^2}. \tag{2}$$

The center x_0 of the beam moves with constant feed velocity u. A cell at site i is

randomly selected and either subject to etching (with probability p) or to diffusion (with probability $1 - p$).

The probability for the particle to be removed by etching is then modelled as the product $p_{ei} = p_{ci} Y_i G_i$. Here $p_{ci} \in [p_0, 1]$ with $0 < p_0 \ll 1$ is the curvature dependent probability for the cell to be removed and corresponds to the curvature of the etching front $\kappa_i = \nabla^2 h_i (1 + (\nabla h_i)^2)^{-3/2}$ at site i taken from an interval $\kappa_i \in [-\kappa_m, \kappa_m]$. The expression $\nabla h_i = (h_{i+1} - h_{i-1})/2a$ denotes the slope and $\nabla^2 h_i = (h_{i-1} - 2h_i + h_{i+1})/a^2$ the discrete Laplacian. Y_i is the absorptivity at site i depending upon the incident angle $\varphi_i = \arctan(\nabla h_i)$. In the model we assumed a function of the form $Y_i = y_0 + y_1 \varphi_i^2 + y_2 \varphi_i^4$. For slopes and curvatures mean values from three adjacents sites were chosen in order to suppress unphysical oscillations, see[7,6] for more details. While the rule for etching had to be modified and deviates
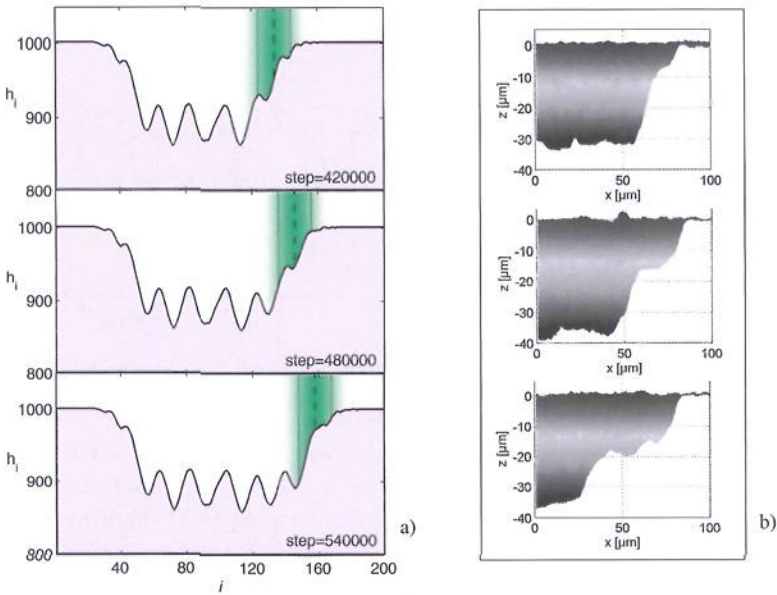
Figure 3. a) Kinetic Monte-Carlo simulations of the discrete model illustrating the evolution of ripple structures in the etching front. b) Comparison with experimentally measured etching fronts.

in several aspects from Cuernos erosion rule, the diffusion rule is adopted from his model. A cell at site i is assumed to move to a randomly chosen nearest neighbour column with a hopping rate

$$w_i^{\pm} = \frac{1}{1 + e^{\beta \Delta \mathcal{H}_{i \to i \pm 1}}} \qquad (3)$$

where

$$\mathcal{H} = J/b^2 \sum_{i=1}^{L-1} (h_i - h_{i+1})^2 \qquad (4)$$

denotes the energy and $\Delta\mathcal{H}_{i\rightarrow i\pm 1}$ is the energy difference between final and initial state of the move, J is a coupling constant and β denotes the inverse temperature[12].

Fig. 3 shows simulations of the discrete model for the following system parameters: $a = 2$, $b = 1$, $\beta J = 5$, $u = 0.002\,a/\text{step}$, $p_0 = 0.15$, $\kappa_m = 0.0825\,b/a^2$. As initial condition a flat surface was chosen. The snapshots in fig. 3a illustrate how ripple structures evolve in the etching front due to the combined action of etching and diffusion. Little troughs in the forefront of the active zone are deepened during the characteristic dwell time ΔT given in eq. (1) while peaks are etched at smaller rates. For comparison, the evolution of an experimentally observed etching front is shown in fig. 3b. For the images the etching process was interrupted at sequential time steps and the surface topography was measured by a 3D-laser-focus scanner[4]. The good correspondence between simulations and experimental results strongly supports the presumption that the basic mechanisms for ripple formation are identified in the discrete model. Fig. 4b shows simulated kerf profiles for different feed
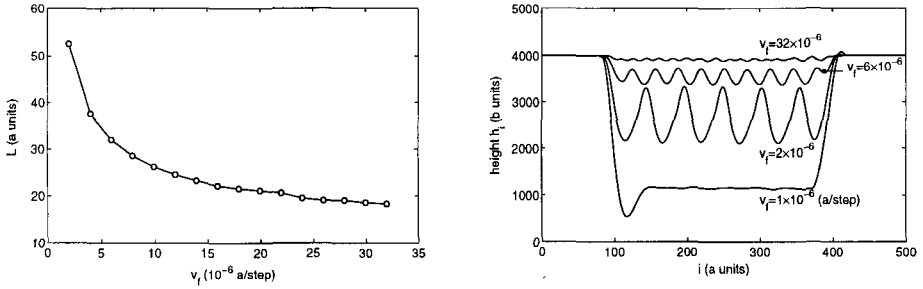
Figure 4. Simulations of the discrete model: a) ripple length vs. feed velocity, b) surface profiles for various feed velocities.

velocities $u = v_f$ of the laser beam. For the lowest velocity $v_f = 1 \times 10^{-6} a/\text{step}$ the etched kerf has the largest depth and shows only some small scale roughness without ripples. When the velocity is doubled ripples with high amplitude are formed spontaneously. For increasing scan speed the amplitude decreases again. For $v_f = 32 \times 10^{-6} a/\text{step}$ the beam is too fast to etch a kerf, i.e. the surface is only scratched. In Fig. 4a the ripple length is plotted as a function of the feed velocity. Like in the experiment, fig. 1, the ripple length decreases for increasing scan speed, i.e. the numerical simulations are qualitatively in good accordance with the experimental observations.

4. The continuous Langevin equation

In this section we summarize in short the steps which are necessary to perform a continuum limit on the basis of the discrete model which is strictly only valid in the small slope limit. A detailed description can be found in[6]. Starting from the rules for etching and diffusion a gain and loss type master equation is set up which

describes the evolution of the joint probability density $P(\mathbf{H}, t)$, i.e. the probability that the surface adopts the configuration $\mathbf{H}$ at time t

$$\frac{\partial P(\mathbf{H}, t)}{\partial t} = \sum_{\mathbf{H}'} W(\mathbf{H}, \mathbf{H}') P(\mathbf{H}', t) - \sum_{\mathbf{H}'} W(\mathbf{H}', \mathbf{H}) P(\mathbf{H}, t) \tag{5}$$

where $W(\mathbf{H}', \mathbf{H})$ is the transition rate per unit time to come from configuration $\mathbf{H}$ to $\mathbf{H}'$. For this purpose the transition probability per unit time τ for erosion and diffusion has to be expressed as a nonlinear function of surface heights $h_i(t)$ and its derivatives $\nabla h_i, \nabla^2 h_i, \ldots$. The master equation is then approximated by a Kramers-Moyal expansion. If the system fluctuations are not too large this expansion can be truncated leading to the associated Fokker-Planck equation, see[16,6]. After regularization and coarse-graining we finally obtain a continuous Langevin equation for the evolution of the etching front. The explicit derivation of this equation needs some calculus and can be found in[6]. Here we give only the result. In a coordinate system comoving with the laser beam with constant velocity u we obtain the continuous Langevin equation which has the form of a modified noisy Kuramoto-Sivashinsky equation which is typical for abrasive techniques[13,14]

$$\frac{\partial h}{\partial t} + u \nabla h = G(x) \left[v_0 + \nu \nabla^2 h - D \nabla^4 h + \right.$$
$$\left. + \lambda (\nabla h)^2 + c_1 \nabla^2 h (\nabla h)^2 + \ldots \right] + \eta(x, t). \tag{6}$$

The advective term on the left hand side arises from the transformation into a comoving coordinate system, $G(x)$ denotes the profile of the active etching zone, v_0 is the mean velocity of the profile in vertical direction in the case of homogeneous etching ($G(x) = 1$). The Laplacian term with coefficient $\nu < 0$ acts like a negative surface tension and results from etching and $D > 0$ arises from thermally activated diffusion. The term $\lambda (\nabla h)^2$ describes a lateral shrinking of the profile which is also due to etching. The nonlinear term with coefficient c_1 has a coarsening effect and $\eta(x, t)$ is a noise term. All coefficients as well as the noise correlations can be expressed explicitly in terms of the parameters of the discrete model.

On the basis of the continuous model equation it is possible to understand the formation of ripples in the etching front. To this end we analyze the linear stability behaviour of the deterministic modified KS equation. Small plane wave perturbations of the form $e^{\omega(k)t + ikx}$ are superimposed to the stationary solution $h_0(x) = C_0 + C_1 x$ of the homogeneous equation and lead to the following dispersion relation

$$\omega(k) = -iku - \nu k^2 - D k^4. \tag{7}$$

For $\nu > 0$ and $D > 0$ the amplitude of the basic solution is always stable. However, for $\nu < 0$ and $D > 0$ plane waves are generated which are unstable for wave numbers $0 < k < k_0 = \sqrt{|\nu|/D}$ and stable for $k > k_0$. The mode with the fastest growing amplitude has the wave number $k_0 = 2\pi \sqrt{2D/|\nu|}$. The behaviour of the noisy KS equation is well studied[12]. In an early stage ripples are formed which are degraded at later stages and replaced by a rough surface structure.

5. Extension to two spatial dimensions and Finite Element simulations

The discrete model can be easily extended to two spatial dimensions x and y. The surface is defined on a 2D-lattice of quadratic cells of size $a \times a$ and represented by height values $h_{i,j}(t)$ at site (i,j), $i = 1, \ldots, L_x$, $j = 1, \ldots, L_y$ measured in units b. As for the 1D model, we assume periodic boundary conditions. For the curvature dependent etching rule one has to assign a unique 'curvature' value $\tilde{\kappa}_{i,j}$ to each lattice point (i,j), for which we propose the following expression

$$\tilde{\kappa}_{i,j} = \frac{\nabla^2 h_{i,j}}{[1 + (\nabla_x h_{i,j})^2 + (\nabla_y h_{i,j})^2]^{\frac{3}{2}}}. \tag{8}$$

Here, $\nabla_x h_{i,j} = (h_{i+1,j} - h_{i-1,j})/2a$ denotes the slope in x-direction, $\nabla_y h_{i,j} = (h_{i,j+1} - h_{i,j-1})/2a$ the slope in y-direction and the Laplacian is approximated as $\nabla^2 h_{i,j} = (h_{i+1,j} + h_{i,j+1} + h_{i,j-1} + h_{i-1,j} - 4h_{i,j})/a^2$. Like in the one-dimensional case, we adopt mean values for $\tilde{\kappa}_{i,j}$ and $\nabla^2 h_{i,j}$ in order to avoid spurious oscillations. The extension of the diffusion rule to the two dimensional case is straightforward. Instead of using eq.(4) the energy has to be determined as follows

$$\mathcal{H} = J/b^2 \sum_{i=1}^{L-1} \left[(h_{i,j} - h_{i+1,j})^2 + (h_{i,j} - h_{i,j+1})^2 \right]. \tag{9}$$

Depending upon the choice of system parameters we obtain as in the one-dimensional case either kerfs with a rough surface or the formation of ripples. In fig. 5a the result of a simulation is presented where the following set of parameter values was chosen: $a = 4$, $b = 1$, $\beta J = 5$, $u = 1.e^{-6}\, a/step$, $p_0 = 0.15$, $\kappa_m = 0.008\, b/a^2$, $\sigma = 10\, a$ and 50% of the laser intensity. The etched kerf clearly shows a ripple structure.

In a next step we want to compare the results of simulations based on the 2D discrete model to numerical solutions of the deterministic noiseless modified KS equation in 2 spatial dimensions using finite elements (FE). To be able to test the FE code, a slightly modified KS equation was used which has been proposed for waterjet cutting[14,5]. This equation differs only in the nonlinear yield function from eq. (6), showing also ripples for certain parameter ranges.

$$\frac{\partial h}{\partial t} = N\, e^{-x_l^2/2\sigma^2} \left(\frac{1}{1 + (\nabla h)^2} + \alpha \nabla^2 h + \beta \nabla^4 h \right). \tag{10}$$

where $x_l = x - ut$ denotes the instantaneous position of the center of the waterjet moving with constant velocity u.

In the following, we briefly summarize the main steps leading to a FE formulation for a parabolic evolution equation of the form

$$\frac{\partial h(\mathbf{x}, t)}{\partial t} = N(\nabla h, \nabla^2 h, \nabla^4 h) \tag{11}$$

where N can be any physically reasonable nonlinear function and the order of the highest derivative is $2m = 4$. For the Galerkin FE method[18,17] eq. (11) is written

in the weak form

$$\mathcal{F} = \int \left[\frac{\partial h(\mathbf{x}, t)}{\partial t} - N(\nabla h, \nabla^2 h, \nabla^4 h) \right] \delta\varphi dx = 0, \qquad (12)$$

where $\delta\varphi$ denotes a test function. A desirable feature of the Galerkin method is that applying Green's theorem m times the order of the highest derivative can be reduced to m, in the case of eq. (12) to $m = 2$. The solution h is approximated by suitable piecewise polynomial trial functions. A necessary condition for convergence to the exact solution requires a complete polynomial of order $m = 2$. In addition, the elements must be conforming which means that the trial functions and their first $m - 1$ derivatives are required to be continuous across element boundaries[18]. In 2D, both conditions are met if we apply two times partial integration to the term with the highest derivative and choose triangular elements with trial and test functions which are complete polynomials of 5th order. For the time integration of the resulting nonlinear system of equations the unconditional stable Crank-Nicolson method is used while nonlinear terms are treated with Newton iterations[17].
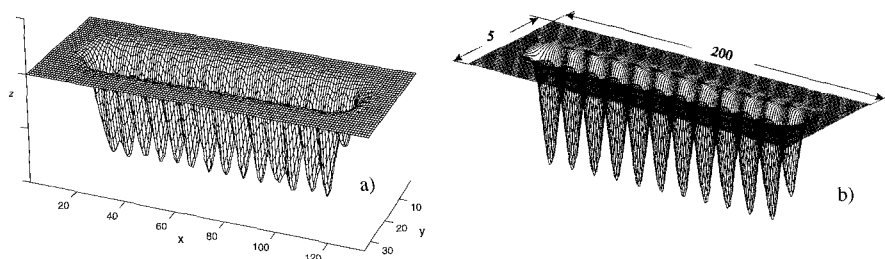
Figure 5.   Profiles of kerfs for increasing feed velocities obtained from (a) discrete model simulations and (b) finite element simulations.

Fig. 5b shows a FE simulation of eq. (10) with a flat surface as initial condition and the system parameters $N = 1.5$, $\alpha = -1.8$, $\beta = -3.283$, $\sigma = 3, u = 0.555$. Regular periodic ripples of wave length $\lambda = 14$ are formed under these conditions. The pattern is in close correspondence to the discrete model simulation for laser-induces etching shown in fig. 5a. It can be concluded that due to the localized ablation process the specific form of the nonlinear terms is not essential for the structure formation. The linear stability analysis has shown that the interplay of the counteracting terms $\nabla^2 h$ and $\nabla^4 h$ is responsible for the formation of periodic patterns.

6. Conclusions

In this study we have presented a simple discrete stochastic model for one and two spatial dimensions for laser induced jet-chemical etching. Numerical simulations have shown that depending upon the system parameters some essential features like

surface roughening and ripple formation can be reproduced with this model. In the continuum limit this model corresponds to a noisy modified Kuramoto-Sivashinsky equation. A linear stability analysis can explain the spontaneous formation of periodic structures. A comparison of 2D simulations of the discrete model and for finite element simulations of a deterministic modified Kuramoto-Sivashinsky for waterjet cutting are in close accordance. In a next step, FE simulations for the deterministic KS equation derived for laser-induced etching, eq. (6), will be compared to discrete model results. Further, refinements of the discrete model are planned like the inclusion of a diffusive layer and the additional heat source which arises in the case of exothermic reactions.

References

1. R. Nowak and S. Metev, A.Phys. A **63**,133 (1996).
2. A. Stephen, G. Sepold, S. Metev and F. Vollertsen, J. Mat. Proc. Tech. **149** 536 (2004).
3. P.T. Pajak, A.K.M. De Silva, J.A. McGeough and D.K.Harrison, J. Mat. Proc. Tech. **149** 512 (2004).
4. T.J. Rabbow, A. Mora, M. Haase and P.J. Plath, to be published in Int. J. Bif. Chaos.
5. G. Radons, T. Ditzinger, R. Friedrich, A. Henning, A. Kouzmichev and E.Westkämper, in: *Nonlinear Dynamics of Production Systems*, eds. G. Radons and R. Neugebauer, WILEY-VCH Verlag Weinheim 575–592 (2004).
6. A. Mora, T. Rabbow, B. Lehle, P.J. Plath and M. Haase, in: *Fractals in Engineering, New trends in theory and applications*, eds. E. Lutton, J. Lévy-Véhel, Springer 125–139 (2005).
7. A. Mora, M. Haase, T. Rabbow and P.J. Plath, http://arxiv.org/abs/cond-mat/0503093 and submitted to Phys. Rev. E (2005).
8. A.-L. Barabási, H.E. Stanley: *Fractal Concepts an Surface Growth*, Cambridge Univ. Press (1995).
9. D.J. Barber, F.C. Frank, M. Moss, J.W. Steeds, I.S.T. Tsong, J. Mater. Sci. **8** 1030 (1973).
10. W. Schulz, V. Kostrykin, M. Niessen, J. Michel, D. Petring, E.W. Kreutz and R. Poprawe, J. Phys. D: Appl. Phys **32** 1219 (1999).
11. I. Georgescu, M. Bestehorn, arXiv:cond-mat/0411244 (2004).
12. R. Cuerno, H.A. Makse, S. Tomassone, S.T. Harrington and H.E. Stanley, Phys. Rev. Lett. **75**, 4464–4467 (1995).
13. K.B. Lauritsen, R. Cuerno, H.A. Makse, Phys. Rev. E **54**, 3577–3580 (1996).
14. R. Friedrich, G. Radons, T. Ditzinger, A. Henning, Phys. Rev. Lett. **85**, 4884–4887 (2000).
15. H. Haken, *Synergetics, Introduction and Advanced Topics*, Springer, Berlin (2004).
16. H. Risken, *The Fokker-Planck equation*, Springer, Berlin (1996).
17. H. Parisch, Lösung der modifizierten Kuramoto-Sivashinsky-Gleichung mit Finiten Elementen am Beispiel des Wasserstrahlschneidens, Internal Report 2005/1, ISD, Universität Stuttgart.
18. C. A. Fletcher, *Computational Galerkin Methods*, Springer, New York (1984).

FRACTAL PROPERTIES OF SOME MACHINED SURFACES

T. R. THOMAS

School of Business and Engineering, Halmstad University, PO Box 823, SE-30118 Halmstad, Sweden

B.-G. ROSÉN

School of Business and Engineering, Halmstad University, PO Box 823, SE-30118 Halmstad, Sweden

Many surface profiles are self-affine fractals defined by fractal dimension D and topothesy Λ. Traditionally these parameters are derived laboriously from the slope and intercept of the profile's structure function. Recently a quicker and more convenient derivation from standard roughness parameters has been suggested. Based on this derivation, it is shown that D and Λ depend on two dimensionless numbers: the ratio of the mean peak spacing to the rms roughness, and the ratio of the mean local peak spacing to the sampling interval. Using this approach, values of D and Λ are calculated for 125 profiles produced by polishing, plateau honing and various single-point machining processes. Different processes are shown to occupy different regions in D-Λ space, and polished surfaces show a relationship between D and Λ which is independent of the surface material.

1 Introduction

Almost as soon as fractal mathematics became available, the possibilities became apparent of its application to various fields of mechanical engineering, including contact mechanics [1,2], heat transfer [3,4], friction and adhesion [5,6] and wear [7,8]. In the related field of manufacturing and production engineering it has been applied to sheet metal rolling and forming [9,10], but a more important area of fractal application is in engineering metrology, specifically the study of surface roughness.

Ever since Whitehouse [11] drew attention to the "parameter rash", the proliferation of ill-defined and often redundant or tautologous parameters used to describe surface roughness, production engineers have been seeking the Holy Grail, a minimum set of parameters, independent of measurement artefacts such as filter cutoffs and sampling interval, which is both necessary and sufficient completely to define the roughness of a surface. One approach to this has been to represent machined surfaces as self-similar fractals, as it was early recognized that the topography of many technical surfaces was fractal in nature. Brown and his co-workers (e.g. [12]) have done much to raise awareness of the possibilities of fractals in the engineering community, and at least one national roughness standard now incorporates fractal dimension. Many proprietary software packages for the characterisation of surface roughness now offer the possibility of calculating the fractal dimension.

However, fractal dimension on its own may not be sufficient. Whereas self-similar fractals are space-filling curves where for any point in the *x-y* plane there may be multiple

values of the argument z, measurements of real surfaces must always be single-valued, that is there can only be one measurement of height for a given point on the surface. This is not to say that the surface itself is necessarily single-valued, simply that real measuring instruments can only show it as such. To preserve the same appearance at any scale, therefore, the surface, as it is magnified on a vertical scale by some numerical factor λ, must be stretched on a horizontal scale by a scaling factor λ^H, where H is the so-called *Hurst exponent* [14] related to the fractal dimension D. Such surfaces are termed *self-affine* [14,15], and their self-affinity is usually characterized by their *topothesy* Λ [13], a parameter with dimensions of length which depends on the Hurst exponent and hence on the fractal dimension.

The fractal dimension D and topothesy are usually obtained from the *structure function* (SF) [14]. The SF is the expected value of the squared difference in heights between points on the surface separated by a delay τ:

$$S(\tau) = <[z(x) - z(x + \tau)]^2>$$

and is closely related to the better-known autocovariance function (ACVF) $R(\tau)$ by

$$S(\tau) = 2[R(0) - R(\tau)] = -\tau^2 R''(0) \tag{1}$$

Then it has been shown [14,15] that

$$S(\tau) = \Lambda^{2D-2} \tau^{2(2-D)} \tag{2}$$

To obtain fractal parameters from Eqn. 2, it has formerly been necessary first to compute the SF from raw profile data and then to obtain the slope and intercept of the logarithmic plot (Figure 1) [15]. This is time-consuming and requires the writing of some code, as no proprietary software currently offers the structure function. As a result, few numerical values of the topothesy of machined surfaces are available in the literature.
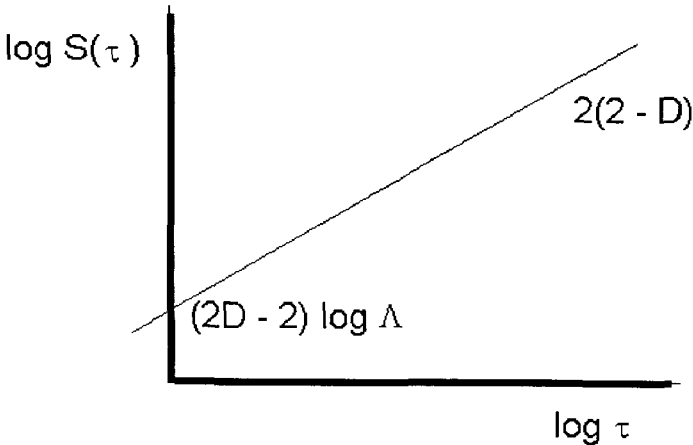
Figure 1. Obtaining D and Λ from the slope and intercept of a logarithmic plot of the structure function [15].

However, Whitehouse [16] has pointed out that the necessary properties of the structure function can be expressed in terms of the derivatives of the ACVF at the origin, which in turn can be obtained from parameters readily available in current surface standards. In this paper we apply Whitehouse's results to some real profiles and give examples of the fractal parameters thus calculated.

2 Derivation of fractal parameters

Whitehouse defines a *curvistructure function* $C(\tau)$ as the expected value of the three-point curvature at a point x:

$$C(\tau) = \langle[2z(x) - z(x+\tau) - z(x-\tau)]^2\rangle$$

By analogy with Eqn. 1,

$$C(\tau) = 2[S(0) - S(\tau)] = \tau^4 R^{IV}(0) \tag{3}$$

where $R^{IV}(0)$ is the 4th derivative of the ACVF at the origin.

Now the local peak density d_p can be expressed in terms of derivatives of the ACVF at the origin (e.g. [15]):

$$d_p = \frac{1}{2\pi}\sqrt{\frac{R^{IV}(0)}{-R''(0)}} \tag{4}$$

With some manipulation of Eqns. 1-4, the fractal dimension can be related to the local peak density and the sampling interval by

$$d_p = \frac{1}{2\tau\pi}\sqrt{4D^2 - 6D + 2} \tag{5}$$

The local peak density is just the reciprocal of the mean local peak spacing S, so finally

$$(2\pi\tau/S)^2 = 4D^2 - 6D + 2 \tag{6}$$

The fractal dimension thus depends only on the dimensionless ratio S/τ (Figure 2). Fractal dimension is high when the local peaks are closely spaced relative to the sampling interval, i.e. the profile contains a substantial short-wavelength component. As the distance between local peaks increases, i.e. there are fewer short-wavelength features, the fractal dimension tends asymptotically to unity.
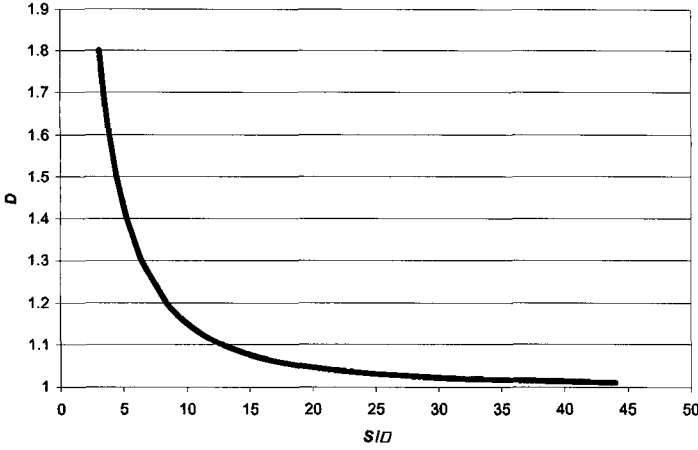
Figure 2. Relationship between fractal dimension D and dimensionless mean local peak spacing S/τ according to Eqn. 6.

Similarly, the zero-crossing density d_0 can be expressed in terms of derivatives of the ACVF:

$$d_0 = \frac{1}{\pi}\sqrt{\frac{-R''(0)}{R(0)}} \tag{7}$$

Combining this with Eqns. 1 and 2, and recalling that R(0) is by definition the square of the rms roughness Rq,

$$\Lambda \;=\; \tau\,(d_0\,Rq\,\pi)^{1/(D-1)}$$

(Whitehouse obtains a slightly different result for Λ). But d_0 is the reciprocal of the mean peak spacing Sm so finally

$$\frac{\Lambda}{\tau} = \left(\frac{\pi Rq}{Sm}\right)^{\frac{1}{D-1}} \tag{8}$$

Thus the dependency of the topothesy on measurable profile properties is described by a relationship between three dimensionless numbers: the fractal dimension, the ratio of the topothesy to the sampling interval, and the ratio of the mean peak spacing to the rms roughness (Figure 3). This last ratio, being the ratio of a texture parameter to an amplitude parameter, may perhaps be thought of as some kind of average reciprocal slope of peaks.
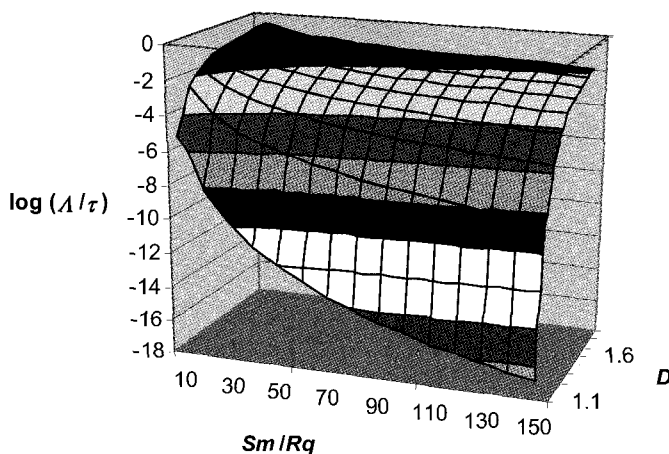
Figure 3.   Relationship between fractal dimension D, dimensionless mean peak spacing Sm/Rq and dimensionless topothesy $Λ/τ$, according to Eqn. 8.

So the estimation of the two fractal parameters reduces to a simple calculation from three existing and widely used roughness parameters, Rq, S and Sm, defined in current technical standards, together with the sampling interval $τ$.   No special computation of any of these parameters should be required as they are implemented in many commercial software packages.

3   Results and discussion

A number of surfaces were selected from a database of 3D height measurements built up over many years.   The original measurements were made on areas of up to 1.5 mm × 1.5 mm of various machined surfaces at sampling intervals of between 0.1 µm and 3.4 µm, using several different models of scanning interferometer.   Finishing processes included polishing, plateau honing, and various single-point machining techniques including milling, turning, reaming and shaping.   Ensemble average values of Rq, S and Sm were obtained from up to 480 parallel profiles on each surface using proprietary software, and from these values fractal dimensions and topothesies were calculated by means of Eqns. 6 and 8 (Figure 4).

Fractal dimensions range from 1.1 to nearly 1.8, a wider range than predicted by Kant [17], who suggested that most real surfaces would have values of D in a narrow range around D=1.5.   The range of values of topothesy is even wider, from ~1 µm for plateau-honed surfaces down to 10^{-20} µm for some of the polished surfaces.   Considered as a length, the latter figure is rather embarrassing, being many orders of magnitude smaller than the diameter of an electron (~10^{-9} µm).   Other workers who have attempted to measure the topothesy have obtained a similarly wide range of results, as reviewed by

Arutyunov [18]. This may arise from Berry's [19] definition of topothesy as the spacing of pairs of points on the surface at which the rms slope will be one radian. Such a slope may be plausible for a geomorphological surface measured on a millimeter scale [18] but is an impossibly high value for any machined surface. Thus, the very low values of topothesy here reported may simply emphasise that some polished surfaces are extremely flat.

It appears that fractal parameters may be of some use to classify finishing processes. Single-point, polished and plateau-honed surfaces seem to occupy separate and reasonably well-defined areas in D-Λ space. Polished surfaces, which include both steel and ceramic surfaces, seem to show a relationship between D and Λ which is independent of the polished material.
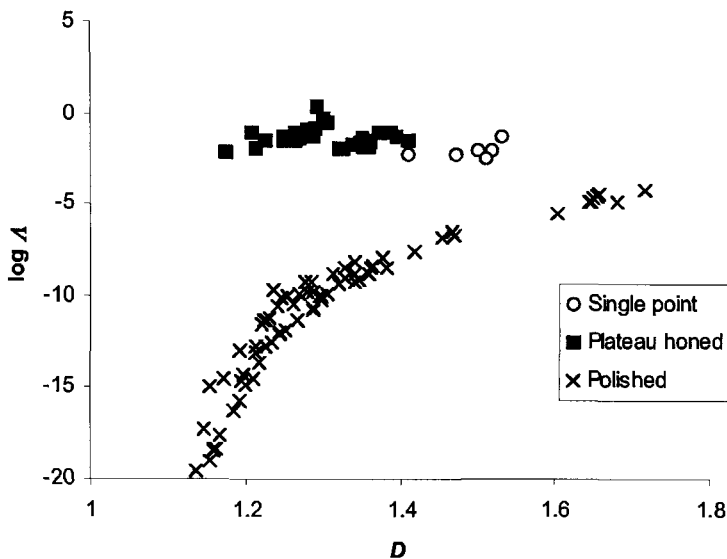
Figure 4. Calculations of fractal dimension and topothesy (in μm) from measurements on 125 profiles.

4   Conclusions

The method proposed by Whitehouse for the calculation of fractal parameters shows a fundamental relationship between these parameters and dimensionless ratios of standard roughness parameters. The method is able to yield a large number of results quickly and conveniently. The numerical values so obtained are self-consistent. In the case of topothesy, this is believed to be the largest collection of values so far presented for technical surfaces, and the numerical values are consistent with values previously reported in the literature. Plots of D against Λ apparently are able to distinguish between different finishing processes. For polished surfaces, there appears to be a relationship between D and Λ which is independent of the polished material.

For workers who do not have access to modern measuring equipment and software, there is an even simpler, though laborious, way of applying this method. If a legacy stylus instrument is available which will measure the average roughness Ra, the rms roughness Rq may be estimated within the limits of measurement uncertainty for the majority of profiles by assuming the sinusoidal relationship $Ra \cong 0.9\ Rq$ [20]. The peak and zero-crossing densities can be counted manually from an analogue recording of the profile itself. In this case the sampling interval should be taken as the low-pass filter cutoff or twice the stylus width, whichever is the greater.

Not all machined surfaces are fractal, and not all fractal surfaces keep the same values of fractal parameters throughout their measured range. Whitehouse [16] has further suggested that simple measurements of the kind described in this paper, if made at two different sampling intervals sufficiently far apart, can quickly establish the fractality of a profile or the existence of multifractality.

All the above arguments apply only to 2D measurements. Extending Whitehouse's method from profiles to surfaces is not straightforward. As we have shown elsewhere [21], both the topothesy and the fractal dimension of profiles on an anisotropic surface vary with the direction of measurement, and it will be for the future to determine how or even whether a 3D topothesy can be defined.

References

1. Bhushan, B. and Majumdar, A., "Elastic-plastic contact model for bifractal surfaces", *Wear* **153**, 53 (1992)
2. Warren, T. L. and Krajcinovic, D., "Fractal models of elastic-perfectly plastic contact of rough surfaces based on the Cantor set", *Int. J. Solids Structures* **32**, 2907 (1995)
3. Majumdar, A., "Effect of interfacial roughness on phonon radiative heat conduction", *Trans. ASME: J. Heat Transfer* **113**, 797 (1991)
4. Wang, S. and Komvopoulos, K., "A fractal theory of the temperature distribution at elastic contacts of fast sliding surfaces", *Trans. ASME: J. Trib.* **117**, 203 (1995)
5. Sahoo, P. and Chowdhury, S. K., "A fractal analysis of adhesion at the contact between rough solids", *Proc. I Mech. E.: J. Eng. Trib.* **210**, 269 (1996)
6. Kennedy, F. E., Brown, C. A., Kolodny, J., Sheldon, B. M., "Fractal analysis of hard disk surface roughness and correlation with static and low-speed friction", *Trans. ASME: J. Trib.* **121**, 968 (1999)
7. Stupak, P. R., Kang, J. H., Donovan, J. A., "Fractal characteristics of rubber wear surfaces as a function of load and velocity", *Wear* **141**, 73 (1990)
8. Tricot, C., Ferland, P., Baran, G., "Fractal analysis of worn surfaces", *Wear* **172**, 127 (1994)
9. Othmani, A. and Kaminsky, C., "Three dimensional fractal analysis of sheet metal surfaces", *Wear* **214**, 147 (1998)
10. Plouraboue, F. and Boehm, M., "Multi-scale roughness transfer in cold metal rolling", *Trib. Int.* **32**, 45 (1999)

11. Whitehouse, D. J., "The parameter rash - is there a cure?", *Wear* **83**, 75 (1982)
12. Brown, C., Johnsen, W. and Hult, K., "Scale-sensitivity, fractal analysis and simulations", *Trans. 7th. Int. Conf. On Metrology & Properties of Engng Surfaces*, pp. 239 (CTH, Göteborg, 1997)
13. Sayles, R. S. and Thomas, T. R., "Surface topography as a nonstationary random process", *Nature* **271**, 431 (1978)
14. Russ, J. C., *Fractal surfaces* (Plenum Press, New York, 1994).
15. Whitehouse, D. J., *Handbook of surface metrology* (Institute of Physics, Bristol, 1994)
16. Whitehouse, D. J., "Some theoretical aspects of structure functions, fractal parameters and related subjects", *Proc. I. Mech. E.* **215**, 207 (2001)
17. Kant, R., "Statistics of approximately self-affine fractals: random corrugated surfaces and time series", *Phys. Rev.* **E53**, 5749 (1996)
18. Arutyunov, P. A., "Fractal analysis of anisotropic surfaces", *Russian Microelectronics* **30**, 411 (2001)
19. Berry, M. V., "Diffractals", *J. Phys.* **A12**, 781 (1979)
20. Thomas, T. R., *Rough surfaces* 2e (Imperial College Press, London, 1999)
21. Thomas, T. R., B.-G. Rosén, N. Amini, "Fractal characterisation of the anisotropy of rough surfaces", *Wear* **232**, 41 (1999)

FRACTALS, MORPHOLOGICAL SPECTRUM AND COMPLEXITY OF INTERFACIAL PATTERNS IN NON-EQUILIBRIUM SOLIDIFICATION*

P. K. GALENKO AND D. M. HERLACH

DLR - German Aerospace Center, Institute of Space Simulation,
Linder Höhe, Cologne 51170, Germany
E-mail: Peter.Galenko@dlr.de

Results of computational modeling are synthesized for the forms of crystals growing in undercooled liquid. Solidification patterns are analyzed as a result of the first-order phase transformation controlled by the heat diffusion, atomic kinetics, and interfacial anisotropy. It is shown that fractal patterns are observed at a vanishing anisotropy of surface energy and atomic kinetics of the solid-liquid interface. Simulated patterns are summarized into morphological spectrum which is considered as a sequence of growth shapes that forms *versus* undercooling (deviation from thermodynamic equilibrium) in solidifying system. A diagram "complexity of growth forms" as a function of "microscopic disorder and deviation from equilibrium" is presented.

1. Introduction

How complex solidification microstructures evolve *versus* deviation from equilibrium is one of the intriguing questions of modern theory of solidification. Answering this question for the concrete system allows one to give systematization of microstructures and find the morphological transitions among them [1]. In systematization of solidification growth-morphologies, a progress has been reached with using experimental, theoretical and computational modeling (see, e.g., monographs [2,3] and collected papers [4]). Recent experimental findings and theoretical description of formation of various growth microstructures are presented in overviews [5].

In addition to understanding of originating of randomly branched patterns in diffusion-limited aggregation [6], qualitative understanding of the development of solidification patterns has been made after introducing interfacial anisotropy of surface energy into the models of crystal growth [7]. Anisotropy provides small singular perturbation on the solid-liquid interface and leads to the selection problem of solidification patterns, especially, to cellular and dendritic patterns (see overviews [8] and refernces therein). The result of interfacial dynamics, leading to the various cellular, fractal and dendritic patterns, has been represented by Ben-Jacob [9] in a

---

form of morphological spectrum dependent on interfacial anisotropy and deviation from equilibrium in Laplacian growth.

In the present paper we consider a spectrum of solidification patterns as a funtion of initial undercooling

$$\Delta = \frac{T_M - T_\infty}{T_Q},\tag{1}$$

where T_M is the melting point temperature (or equilibrium temperature of solidification); T_∞ is the temperature of the liquid phase far from the interface; T_Q is the unit of undercooling (adiabatic temperature of solidification) defined by relation $T_Q = Q/\kappa$, Q and κ are the latent heat of solidification and specific heat per unit volume, respectively. Dimensionless undercooling Δ may be considered as a parameter characterising deviation from thermodynamic equilibrium. It is chosen as the main parameter describing morphological changes of crystal patterns. Therefore, we present results of computational modeling of solidifying patterns depending on undercooling and at a given fixed anisotropy of interfacial properties such as interfacial energy and interfacial kinetics of crystal growth. A transition to fractals in this morphological spectrum is shown as a result of vanishing anisotropy of interfacial properties. Finally, we analyze complexity of interfacial patterns (fractals and dendrites) and a transition to homogeneous (non-branched) patterns *versus* disorder in a solidifying system and deviation from thermodynamic equilibrium.

2. The model

In order to obtain the morphological spectrum of patterns, we have used the phase-field model for solidification. In contrast to the classical sharp-interface model in which one assumes the zero thickness of the phase interface, the phase-field model introduces a new variable, the phase-field Φ, and replaces the sharp interface by a diffuse interface. The phase-field Φ has a constant value in phases, e.g. $\Phi = 1$ for initially unstable phase which is transforming into the final phase with $\Phi = -1$. Between these phases, in the interfacial region, the phase field, Φ, changes steeply but smoothly from -1 to 1. It allows one to avoid the explicit interface tracking and locate the interface at $\Phi = 0$. Consequently, the phase-field Φ is considered as an order parameter which is introduced to describe the moving interfacial boundary between initially unstable phase and final phase.

We use the phase-field model via "thin-interface" analysis developed by Karma and Rappel [10,11] and extended by Bragard et al. [12] for both small and highest undercooling in a pure system. In this analysis the interface thickness W is assumed small compared to the scale of the crystal pattern but not smaller than the microscopic capillary length d_0.

The main governing equations of the phase-field model are written as
- *energy conservation*

$$\frac{\partial T}{\partial t} = a\nabla^2 T + \frac{T_Q}{2}\frac{\partial \Phi}{\partial t} - T_Q \nabla \cdot \vec{q},\tag{2}$$

- phase-field evolution

$$\tau(\vec{n})\frac{\partial \Phi}{\partial t} = \nabla \cdot \left(W^2(\vec{n})\nabla \Phi\right)$$

$$+ \sum_{\omega = x,y,z} \frac{\partial}{\partial \omega}\left(|\nabla \Phi|^2 W(\vec{n})\frac{\partial W(\vec{n})}{\partial(\partial_\omega \Phi)}\right) - \frac{\partial F}{\partial \Phi} + \theta(\vec{r},t), \tag{3}$$

where T is the temperature, $\vec{r}$ is the radius-vector of a point with the Cartesian coordinates (x,y,z), t is the time, a is the thermal diffusivity, and the phenomenological free energy F is defined by

$$F(T,\Phi) = f(\Phi) + \frac{\lambda(T - T_M)}{T_Q}g(\Phi). \tag{4}$$

Eq. (4) includes the double-well function $f(\Phi) = -\Phi^2/2 + \Phi^4/4$ and the odd function $g(\Phi) = \Phi - 2\Phi^3/3 + \Phi^5/5$ itself, and the form of free energy F is constructed in such a way that a tilt λ of an energetic well controls the coupling between the fields of T and Φ.

The time $\tau(\vec{n})$ of the phase-field kinetics and the thickness $W(\vec{n})$ of the anisotropic interface are given by

$$\tau(\vec{n}) = \tau_0 a_c(\vec{n})a_k(\vec{n})\left[1 + a_2\frac{\lambda d_0}{a\beta_0}\frac{a_c(\vec{n})}{a_k(\vec{n})}\right],$$
$$W(\vec{n}) = W_0 a_c(\vec{n}), \tag{5}$$

where τ_0 is the time-scale for the phase-field kinetics, W_0 is the parameter of the interface thickness, d_0 is the microscopic capillary length, and $\vec{n}$ is the normal vector to the interface. The second term in square brackets of Eq. (5) for $\tau(\vec{n})$ defines a correction for the "thin-interface" asymptotics with adopting a constant $a_2 = 0.6267$ [12] in comparison with the sharp-interface analysis. The functions $a_c(\vec{n})$ and $a_k(\vec{n})$ of interfacial anisotropy are described by
- the anisotropy of interfacial energy

$$a_c(\vec{n}) \equiv \frac{\gamma(\vec{n})}{\gamma_0} = (1 - 3\varepsilon_c)\left[1 + \frac{4\varepsilon_c}{1 - 3\varepsilon_c}(n_x^4 + n_y^4 + n_z^4)\right], \tag{6}$$

where $\gamma(\vec{n})$ is the surface energy dependent on the normal vector $\vec{n}$ to the interface, γ_0 is the mean value of the interfacial energy along the interface, and ε_c is the anisotropy parameter,
- the anisotropy of kinetics of atomic attachment to the interface

$$a_k(\vec{n}) \equiv \frac{\beta(\vec{n})}{\beta_0} = (1 + 3\varepsilon_k)\left[1 - \frac{4\varepsilon_k}{1 + 3\varepsilon_k}(n_x^4 + n_y^4 + n_z^4)\right], \tag{7}$$

where $\beta(\vec{n})$ is the kinetic coefficient dependent on the normal vector $\vec{n}$ to the interface, β_0 is the averaged kinetic coefficient along the interface, and ε_k is the kinetic

anisotropy parameter. In Eqs. (5)-(7), the normal vector $\vec{n} = \vec{n}(n_x, n_y, n_z)$ has the components defined by the gradients of the phase-field as follows

$$n_x^4 + n_y^4 + n_z^4 = \frac{(\partial\Phi/\partial x)^4 + (\partial\Phi/\partial y)^4 + (\partial\Phi/\partial z)^4}{|\nabla\Phi|^4}. \tag{8}$$
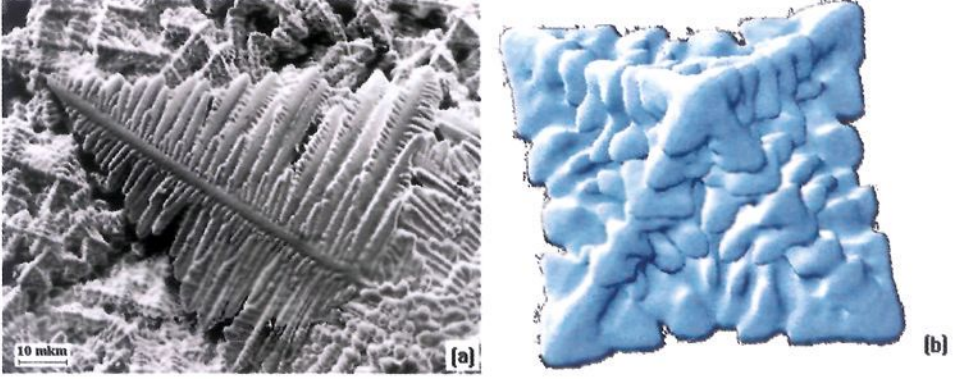
Figure 1. (a) Realistic dendrite grown from undercooled metallic melt and visualized with using electronic microscope upon solidification. (b) 3D modeled dendritic pattern of nickel crystal grown from undercooled melt in the phase-field model (2)-(10).

As it has been theoretically proved by Brener and Temkin and experimentally verified by Bisang and Bilgram [14] thermal fluctuations are necessary to be taken into account for observed realistic shape of crystals with the interface sidebranching. To model the branched patterns, the random variables $\vec{q}(\vec{r}, t)$ and $\theta(\vec{r}, t)$ for description of the thermodynamic noise existing in solidifying systems have been introduced into equations (2) and (3), respectively. The variables were chosen in consistency with the results of Refs. [11,13] in which analysis of stochastic forces is developed for the anisotropic phase-field models. According to Karma and Rappel [11], the noise is added by introducing the non-conserved random variable $\theta(\vec{r}, t)$ for the phase-field and the vector $\vec{q}(\vec{r}, t)$ of conserved thermal noise to be uncorrelated in space and time and obeying Gaussian distributions. The variables have the variances dictated by the fluctuation-dissipation theorem which are described by

$$\langle \theta(\vec{r}, t)\theta(\vec{r}\,', t) \rangle = 2\lambda d_0^3 F_{exp}\delta(\vec{r} - \vec{r}\,')\delta(t - t\,'),$$

$$\langle q_m(\vec{r}, t)q_n(\vec{r}\,', t) \rangle = 2D d_0^3 F_{exp}\delta_{mn}\delta(\vec{r} - \vec{r}\,')\delta(t - t\,'). \tag{9}$$

In Eq. (9), F_{exp} is a single dimensionless experimental parameter which represents the magnitude of the noise existing in experiments and is defined by the following relation

$$F_{exp} = \frac{k_B T_M^2}{T_Q Q d_0^3}, \tag{10}$$

where k_B is the Boltzmann constant.

3. Morphological spectrum

The phase-field model (2)-(10) allows one to model quantitatively the dendritic growth and morphological transformations among the growth forms with the acceptance of material constants obtained experimentally [15]. This system has been resolved numerically on the uniform numerical grid with application of the explicit numerical scheme for simultaneous solution of the heat equation (2) and phase-field equation (3). Initial and boundary conditions, the numerical parameters and constants of material applicable to pure nickel are described in Ref. [15].
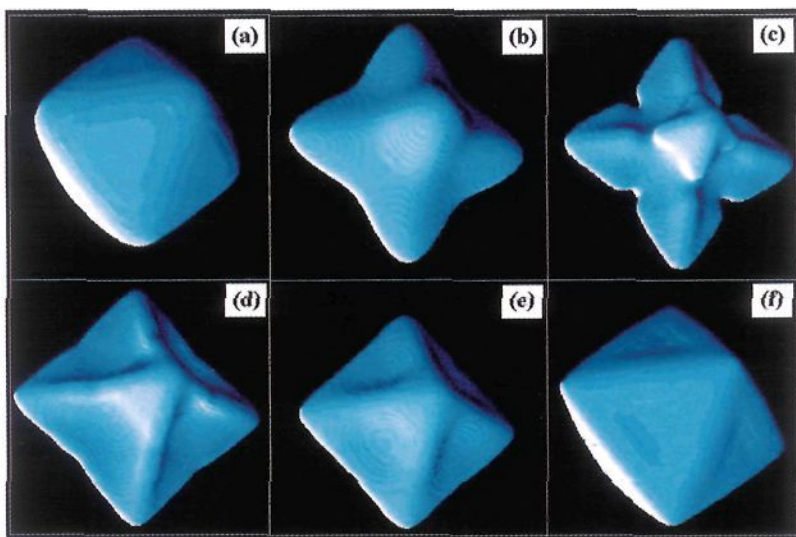
Figure 2. Simulated morphologies for different initial undercoolings: (a) $\Delta = 0.15$; (b) $\Delta = 0.25$; (c) $\Delta = 0.50$; (d) $\Delta = 0.75$; (e) $\Delta = 0.85$; (f) $\Delta = 1.01$. The patterns were modeled on a numerical grid of size 450^3.

Phase-field modeling of crystal growth in liquids showed good reproduction of crystalline structure in comparison with a structure of realistic dendrites grown from the metallic melt. Figure 1 represents the result of modeled dendrite grown from nickel melt in comparison with dendritic pattern observed in electronic microscopy upon solidification of undercooled melt. It can be seen that the phase field model (2)-(10) gives realistic results of modeling in comparison with experimental observations. One of the interesting features of dendritic growth may be outlined from this computation: the secondary branches on the main stems of the modeled dendrite (1(b)) are the result of stochastic noise described by Eqs. (9)-(10). The main stems are result of the existence of the preferable crystallographic directions for growth (for nickel crystals, these give the four-fold symmetry) at a given undercooling. In contrast, the surfaces of the main developed stems become morphologically unstable and give a secondary branches in a presence of stochastic noise.

In the numerical simulations, we obtained a morphological spectrum of different interfacial crystal structures for a range of dimensionless undercoolings $\Delta = 0.15, \ldots, 1.1$. The results are presented in Fig. 2 and show a transition from smooth non-branched crystals at small undercoolings $\Delta \leq 0.15$ to dendritic patterns at intermediate undercoolings $0.15 < \Delta < 1.0$ changing again to non-branched crystals at high undercoolings $\Delta \geq 1.01$. Within the range of intermediate undercoolings, the shape of the dendrites is dictated by preferred crystallographic directions. For the case of nickel, the dendrites grow in the $< 100 >$-direction even though both physical anisotropies given by Eqs. (6)-(7) act during growth simultaneously.

The role of anisotropy for selecting and stabilizing dendritic structures has been reviewed in Refs. [8]. In the presence of both, surface energy and kinetic anisotropy, a stable growth mode of dendrites is found for arbitrary undercoolings. As classified in Refs. [8,16], the so-called surface tension needle crystals occur at small and moderate undercoolings. At higher undercoolings, kinetic needle crystals grow due to the anisotropy of the atomic kinetics. As soon as one of the anisotropies vanishes, one may expect a degeneration of the stable dendritic growth within the considered range of undercoolings.
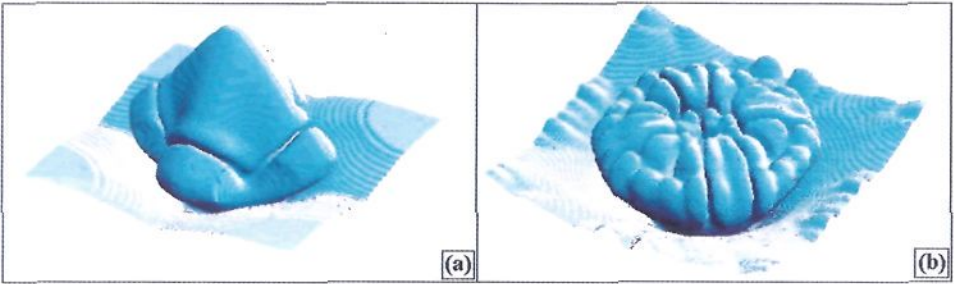
Figure 3. Dendritic and fractal patterns in three-dimensional phase-field modeling. Simulated morphologies of crystals are obtained for: (a) dendrite at $\Delta = 0.95$ and with finite values of anisotropy strengths ε_c and ε_k ; (b) fractal at $\Delta = 0.95$ and with zero values of anisotropy strengths: $\varepsilon_c = \varepsilon_k = 0$ in Eqs. (6)-(7). Patterns were modeled on a numerical grid of size 550^3.

4. Fractals in crystal growth

A transition from regular dendritic patterns to fractals is observed for vanishing anisotropy of crystal growth. To give evidence of this idea, we simulated a growth of a planar interface without stochastic noise but with a small perturbation on it (as an initial seed for instability). The results of modeling for growth both with and without crystalline anisotropy are presented in Fig. 3. As shown in Fig. 3(b), the branched crystals of fractal type develop for zero anisotropies of the surface energy $\epsilon_c = 0$ and of the kinetic coefficient $\epsilon_k = 0$ in Eqs. (6)-(7).

Due to the absence of crystalline anisotropy, tips of branches split and the new branches grow up to their own instability again. As a result of this splitting, so-
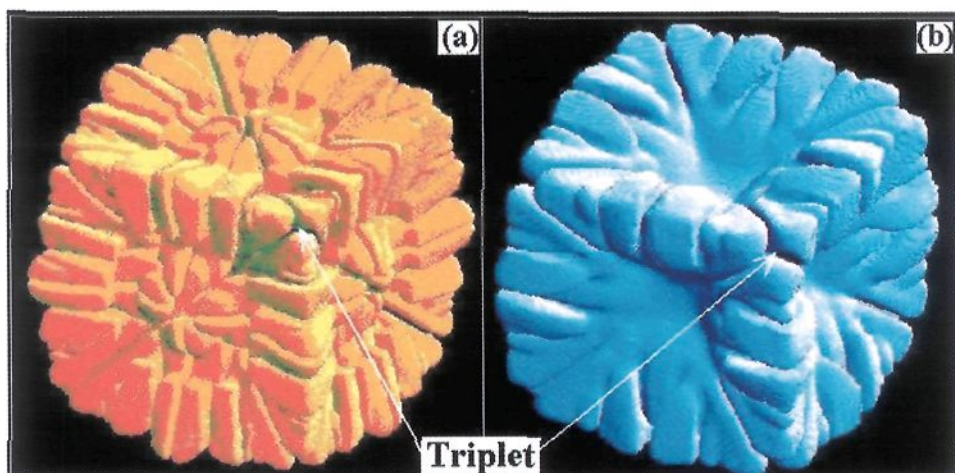
Figure 4. Fractal patterns observed for zero anisotropies $\epsilon_c = \epsilon_k = 0$ on a numerical grid of size 550^3. The formation of fractals was found at the undercooling: (a) $\Delta = 0.70$ and (b) $\Delta = 0.80$.

called "dublons" in 2D and so-called "triplets" in 3D appear at the tips of developed branches. A triplet consisting of three cooperating symmetry-broken fingertips formed at the center of the fractal-like structure is shown in Fig. 4. Such triplets were first modelled as dynamically stable objects in a 3D space using the phase-field method in Ref. [17]. We observed the formation of triplets in liquids at the undercoolings $\Delta < 1.0$ as is shown in Fig. 4. Beyond the hypercooling limit of $\Delta > 1.0$, smooth spherical structures grow as crystals without any anisotropy [15].

5. Complexity in crystal growth

Considering the morphological spectrum, Fig. 2, one may outline one interesting feature: patterns which solidify at very small undercooling ($\Delta T < 0.15$) and very large undercooling ($\Delta T > 1.0$) exhibit smooth non-branched forms [a]. In the range of moderate undercoolings, crystals have branched and complicated forms with secondary branches developed along preferable crystallographic directions for growth.

In the present description, we consider complexity in crystal growth as a complexity in branched patterns *versus* disorder in solidifying system [18]. Two informal definitions may be drawn. We define:

- complexity as a measure of structural inhomogeneity in the system;
- disorder as a parameter characterizing atomistic kinetics around the solid-liquid interface.

---

[a]It is true, certainly, for a chosen size of the computational domains used in the present work. For computational domains larger than presently used ones, one may expect a developing of morphological instability for crystals grown at very small undercoolings (e.g., for undercooling $\Delta T < 0.15$, Fig. 2(a)). After such morphological instability, the crystals becomes branched ones.
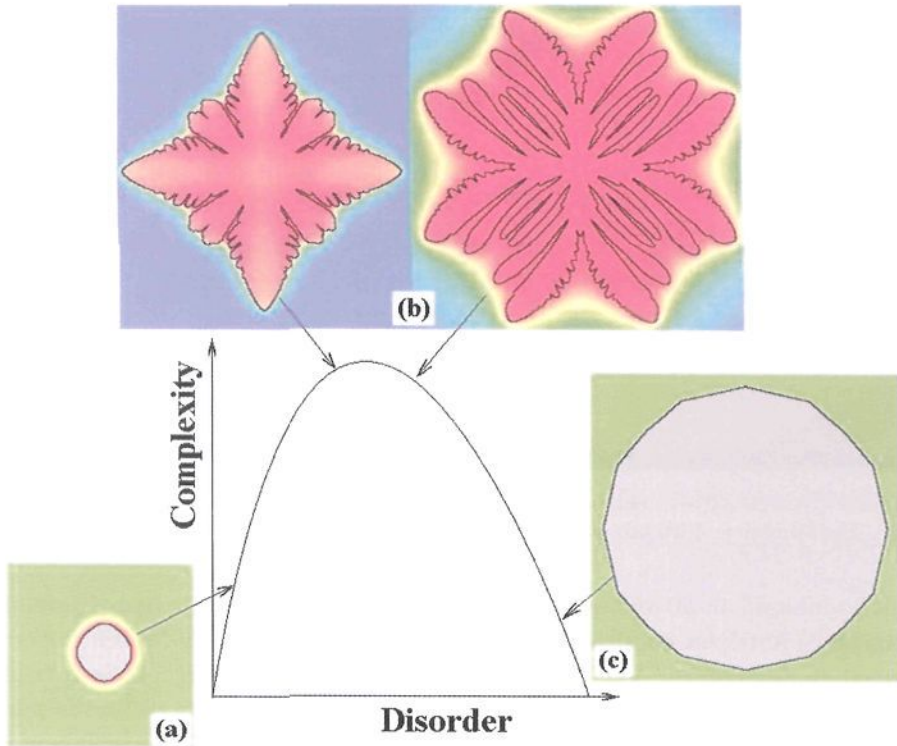
Figure 5. Generic complexity of crystal patterns *versus* disorder in solidifying system. (a) Small deviations from equilibrium (a region of very small undercooling) and low kinetics of atoms attachment to the solid-liquid interface lead to simplicity of the crystal form. (b) Moderate deviations from equilibrium. Dendrite (left) is formed due to existence of anisotropy of interfacial properties. Fractal (right) is developed as a crystal with absence of interfacial anisotropy. (c) Large deviations from equilibrium (a region of very high undercooling) and very intensive kinetics of atoms attachment to the solid-liquid interface lead to simplicity of the crystal form.

From these definitions, one may expect to observe smooth and uniform forms of crystals for both small disorder and high disorder in atomistic kinetics. Complex crystals can be some place "in the middle". Figure 5 demonstrates this idea [b].

With small deviations from equilibrium (small undercooling), the driving force for atomistic attachment to the solid-liquid interface is very small, and the diffusion process ahead of the interface has enough time for smoothing of inhomogeneities in the thermal diffusion field. The shape is morphologically stable and it has very simple form without branches. This result is demonstrated in Fig. 5(a) for 2D modeling and it can be compared with the Fig. 2(a) for 3D modeling of crystal growth at small undercooling of the liquid. Thus, a high order on the atomistic

---

[b]Such types of relationships "complexity - disorder" to appear in the literature seems to have been the similar diagrams "complexity - entropy" given by J.P. Crutchfield and K. Young, *Phys. Rev. Lett.* **63**, 105 (1989).

level (low values of disorder) leads to smooth and simple crystal forms on the mesoscopic/macroscopic level of description.

With moderate deviations from equilibrium (region of normal undercooling), the intensity of atomistic attachment to the solid-liquid interface does not provide stabilizing effect on the preserving of growth form against temperature gradients ahead of the interface in the undercooled liquid. The shape becomes morphologically unstable and the secondary branches appear. This result (applicable to both dendrites and fractals) is demonstrated in Fig. 5(b) for 2D modeling. It can be compared with the Fig. 2(b)-(e) for 3D modeling of crystal growth in undercooled liquid as well. Thus, a moderate disorder on the atomistic level leads to complexity in growth forms on the mesoscopic/macroscopic level of description of pattern formation.

With large deviations from equilibrium (very large undercooling), the driving force for atomistic attachment to the solid-liquid interface is very high. It leads to smoothing of microscopical inhomogeneities of the solid-liquid interface. The crystal shape is morphologically stable and it has very simple form without branches. This result is demonstrated in Fig. 5(c) for 2D modeling and it can be compared with the Fig. 2(f) for 3D modeling of crystal growth in highly undercooled liquid. A transition to smooth patterns at higher undercoolings is known as a globular transition in crystal growth [3,12]. Thus, high intensity of atomistic kinetics on the microscopic level leads to smooth non-branched crystal forms on the mesoscopic/macroscopic level of description.

Consequently, Figure 5 has a conceptual role in description of pattern formation. It shows that at low and high disorder on the microscopic level (directly connected with atomistic kinetics) the patterns have low complexity in their own mesoscopic/macroscopic growth forms. In contrast, at the mid of microscopic disorder (provided by moderate deviations from thermodynamic equilibrium) the developing patterns exhibit variety and complexity.

6. Conclusions

Using computational modeling, non-equilibrium solidification of undercooled liquid has been analyzed for synthesizing results of the crystal growth forms. The role of anisotropy of the interface for selecting crystal patterns has been demonstrated in example of dendritic and fractal crystals. Particularly, it has been shown that the fractal crystals are patterns which are growing with isotropic surface energy and kinetics of atomic attachment to the interface.

The morphological spectrum of crystals was analyzed as a sequence of forms in dependence from undercooling of a liquid (initial deviation from thermodynamic equilibrium). It was shown that patterns which solidify at very small undercooling and very large undercooling exhibit smooth non-branched forms. The branched patterns are in the intermediate region of undercooling. This fact has been used to construct a diagram "complexity-disorder" for the solidifying patterns. The diagram

has a conceptual role in analysing mesoscopic/macroscopic growth forms *versus* disorder on the microscopic level which is directly correlated with atomistic kinetics at the interface. The maximum of complexity in growth forms has been observed at the intermediate deviation from equilibrium and a middle level of intensity of atomic kinetics.

References

1. P.K. Galenko and M.D. Kirivilyov, In: *Fractal Frontiers*. Edited by M.M. Novak and T.G. Dewey (World Scientific, Singapore, 1997), p.411; *Modelling Simul. Mat. Sci. Eng.* **8**, 81 (2000).
2. T. Vicsek, *Fractal Growth Phenomena*, 2nd Edition. (Singapore: World Scientific, 1992).
3. P.K. Galenko and V.A. Zhuravlev, *Physics of Dendrites* (World Scientific, Singapore, 1994).
4. *Growth and Form*, ed. M. Ben Amar et al. (Plenum Press, New York, 1991).
5. W.J. Boettinger, S.R. Coriell, A.L. Greer, A. Karma, W. Kurz, M. Rappaz and R. Trivedi, *Acta Mater.* **48**, 43 (2000); D.M. Herlach, *J. Phys.: Condens. Matter* **13**, 7737 (2001); H. Müller-Krumbhaar, Th. Abel, E. Brener, M. Hartmann, N. Eissfeldt and D. Temkin, *JSME Int. Journal* **B45**(1), 129 (2002).
6. T.A. Witten and L.M. Sander, *Phys. Rev Lett.* **47**, 1400 (1981); T. Vicsek, *Phys. Rev. Lett.* **53**, 2281 (1984).
7. D.A. Kessler, J. Koplik and H. Levine, *Phys. Rev.* **A30**, 3161 (1984); E. Ben-Jacob, N.D. Goldenfeld, B.G. Kotliar and J.S. Langer, *Phys. Rev. Lett.* **53**, 2110 (1984).
8. D. Kessler, J. Koplik and H. Levine, *Adv. Phys.* **37**, 255 (1988); E. Brener and V.I. Melnikov, *Adv. Phys.* **40**, 53 (1991).
9. E. Ben-Jacob, *Contemp. Phys.* **34**, 247 (1993).
10. A. Karma and W.-J. Rappel, *Phys. Rev. E* **E57** (1998) 4323.
11. A. Karma and W.-J. Rappel, *Phys. Rev.* **E60** (1999) 3614.
12. J. Bragard, A. Karma, Y. H. Lee and M. Plapp, *Interface Science* **10**(2-3) (2002) 121.
13. S.G. Pavlik and R.F. Sekerka, *Physica* **A286**, 283 (1999).
14. E. Brener and D. Temkin, *Phys. Rev.* **E51**, 351 (1995); U. Bisang and J.H. Bilgram, *Phys. Rev. Lett.* **75**, 3898 (1995).
15. B. Nestler, D. Danilov and P. Galenko, *J. Comp. Physics* **207**, 221 (2005).
16. F. Liu and N. Goldenfeld, *Phys. Rev.* **A42**, 5052 (1990).
17. T. Abel, E. Brener and H. Müller-Krumbhaar, *Phys. Rev.* **E55**, 7789 (1996).
18. D. Herlach, D. Holland-Moritz, P. Galenko and A. Wiek, *Selbstorganisation und Strukturbildung. Lecture notes in Bochum University.* (Ruhr-Universität Bochum, Fakutlät für Physik und Astronomie, 2005).

COMPETITION OF DOUBLON STRUCTURE IN THE PHASE-FIELD MODEL

SEIJI TOKUNAGA

Department of Applied Science for Electronics and Materials, Kyushu Univ., Kasuga-Koen 6-1, Kasuga, Fukuoka,
816-8580, Japan
Email: tokuna1@asem.kyushu-u.ac.jp

HIDETSUGU SAKAGUCHI

Department of Engineering Sciences for Electronics and Materials, Kyushu Univ., Kasuga-Koen 6-1, Kasuga,
Fukuoka, 816-8580, Japan

A doublon is one of the fascinating patterns found in crystal growth. It forms a pair of symmetry broken fingers. In this paper, we investigate the doublon pattern growing with the effect of other doublons. We have found that as the result of some competition, some doublons break their own structure and the number of doublon keeping its structure depends on diffusion length.

1. Introduction

The snowflake is a typical dendritic crystal of ice. The dendrite has been intensively studied, and it was found that the crystal anisotropy is essentially important for the steady growth of dendrites [1-3].
    A doublon is a noteworthy pattern found by Ben Amar and Brener, in that it is a pattern growing steadily even in the isotropic system and is discovered in the field of crystal growth. It is a pair of symmetry broken fingers and has a stable groove. There are many fascinating researches of doublon. Akamatsu *et al.* found doublon patterns in an experiment of directional solidification [4]. They predicted that doublon pattern exists in experimental condition of low anisotropy and high supercooling. Losert *et al.* investigated the stability of doublon structures in an experiment of directional solidification and a numerical simulation with a phase-field model [5]. Samid-Merzel *et al.* found doublon patterns in an experiment of drying water film [6]. Recently, doublon-like pattern was found in diffusion-limited growth of xenon crystals [7].
    A phase-field model is a useful simulation model for the crystal growth and studied by many authors [8-10]. It has a form of the Ginzburg-Landau equation coupled with the diffusion equation. For the melt growth, Ginzburg-Landau equation represents the dynamics of phase transition from the liquid phase to the solid one. Karma and Rappel improved the phase-field model to simulate the case of negligible interface kinetics that is physically relevant at the low supercooling for a large class of materials [11].
    We used phase-field model to investigate the doublon pattern. We classified the doublon into surface tension doublon and kinetic one. Both are formed from small seeds of crystal of different sizes in wide parameter range [12]. A doublon can be formed through a tip-splitting process from a small single seed of crystals [13]. We also found a doublon pattern whose groove is oscillating. The groove oscillation does not decay in a narrow parameter range [14].
    The doublon growing with the competition of any other doublons has not been investigated though a lot of researches of doublon growing up alone have been published up to now. In this paper, we report the growth of doublon pattern under the competition of group of doublons using the phase-field model.
    In Sec.2, we introduce the model equations of the phase-field model. In Sec.3, the detail of numerical method used in this simulation is explained. In Sec.4, the results of simulation are shown and discussed. In Appendix, we explain a discretization method especially valid for isotropic case.

2. Model equation

The model equation neglecting the crystal anisotropy of the phase-field model is expressed as

$$\tau\partial_t p = \{p - \lambda u(1 - p^2)\}(1 - p^2) + W\nabla^2 p \tag{1}$$

$$\partial_t u = D\nabla^2 u + \partial_t p/2, \tag{2}$$

where, λ is a dimensionless parameter that controls the strength of the coupling between the phase and diffusion fields, p is an order parameter and $p=1$ and $p=-1$ correspond to respectively, solid and liquid phase, τ and W are a time constant and diffusion constant, respectively. The variable u is the dimensionless temperature that is expressed as $u=(T-T_M)/(L/C_P)$, where T, T_M, L and C_p are respectively, the temperature, the melting temperature, the latent heat and the specific heat at constant pressure. The diffusion constant for u is denoted by D. The term $\partial_t p/2$ in Eq. (2) represents latent heat production at the interface.

The isotropic version of the fundamental equation of crystal growth that Karma and Rappel derived as the sharp-interface limit of the phase-field model is expressed as

$$\partial_t u = D\nabla^2 u, \tag{3}$$

$$u_i = -d_0\kappa - \beta\upsilon_n. \tag{4}$$

Equation (3) is the diffusion equation for u and Eq. (4) is the generalized Gibbs-Thomson condition, respectively, where d_0, κ, β and v_n denote respectively, the isotropic capillary length, the interface curvature, the isotropic kinetic coefficient and the normal interface velocity. These parameters are expressed using W and τ as

$$d_0 = \frac{I}{\lambda J} \tag{5}$$

$$\beta = \frac{I}{\lambda J}\left[1 - \lambda\frac{1}{2D}\frac{K + JF}{I}\right], \tag{6}$$

where, $I = 2\sqrt{2}/3$, $J = 16/15$, $F = \sqrt{2}\ln 2$, $K = 0.13604$. If the parameter λ is chosen as $\lambda = (2ID)/(K + JF)$, the isotropic kinetic coefficient vanishes, that is, the kinetic effect becomes negligible. This is the advantage of the phase-field model compared with other models. In this paper, we neglect the kinetic effect.

3. Simulation

The simulations are performed in a square box of size $L_x \times L_y = L_x \times 480$. We simulate in the case of $L_x=480$ and 320 to investigate whether the stability of the doublon depends on the lattice size or not. It has a periodic boundary condition with respect to horizontal direction, i.e. $p(0,y,t)=p(L_x,y,t)$, $u(0,y,t)=u(L_x,y,t)$, and has a mirror symmetric boundary condition with respect to vertical direction. In initial condition, 10 seeds of crystal of radius 2.4, are set at random x ($y=0$) positions. All the space except for the seeds of crystal is initially supercooled Δ. If the tip of the pattern reaches $y=440$, the simulation is stopped, and the numerical data for the order parameter and the temperature are saved in our computer. The order parameter and the temperature profiles in the tip region $(p'(x,y),u'(x,y))$ are used for the initial conditions for the next step, that is, $p(x,y,t=0)=p'(x,y+320)$, $u(x,y,t=0)=u'(x,y+320)$ for $y<160$, and $p(x,y,t=0)=-1$, $u(x,y,t=0)=-\Delta$ for $y>160$. By iterating this process, we can observe the growth of doublons for a long time. We set $W=\tau=1$ to make the simulation easily in the isotropic case.

4. Result and discussion

Figures 1 show the group of doublon patterns at $\Delta=0.74$ and $L_x=480$. Figure 1(a) displays the first simulation from 10 crystal seeds at random positions. At first, a growing finger is affected by a nearest neighboring finger and make a pair, after that, the doublon pattern is made. A finger, which cannot make a pair, isn't able to grow up because other doublons growing faster than one finger would consume the diffusion field, which is necessary to grow up. The nearer initial arrangement of two crystals is, the quicker it grows, because a doublon is formed early. The velocity of all the doublons $v=0.165$. So the diffusion length, which is the effective length of diffusion field, $l_d=2D/v=24.2$. The left doublon is

curving and approaching to the central doublon. As a result, a finger of the left doublon can't grow and the doublon structure is certainly broken. At the end of the simulation, two doublons survived. Figure 1(b) is second simulation started from the result of first simulation. The central doublon breaks its symmetric pattern, however, two fingers of doublon grow together and keep their stability. Figure 1(c), (d) and (e) are third, fourth and fifth simulations, respectively. The right doublon gradually approaches to taller left doublon and the structure of right doublon is gradually broken as time passes. Figure 1(f) is the sixth simulation. The right doublon approaches to the left doublon more and more. After that, the structure of right doublon is perfectly broken. On the contrary, the left doublon keeps its doublon structure even though symmetrical form is partly broken affected by the right doublon. Figure 1(g) is the seventh simulation. The surviving doublon recovers its symmetrical form and grows alone. Figure 1(h) is the eights simulation. Only one doublon grows keeping its stability like it has grown alone from the beginning.

Figures 2 show the group of doublon patterns at Δ =0.76 and L_x=480. Figure 2(a) is first simulation. After the competition, three doublons retain their structure. Figure 2(b) and (c) are second and third simulations, respectively. In Fig.2(b), the left doublon approaches to central doublon, and its symmetry is partly collapsed. And then, the left doublon in Fig. 2(c) is collapsed by the effect of central doublon and cannot be observed. Figure 2 (d), (e), (f), (g) and (h) are fourth, fifth, sixth, seventh and eighth simulation respectively. These two doublons are growing together and keep their stability for a long time. Finally, two doublons remain with these parameters Δ =0.76, L_x=480.

Ben Amar and Brener calculated that the tip velocity of doublon is in proportion to 9th power of supercooling Δ [15]. We also confirmed the same power law of doublon velocity in the phase-field model [12]. Therefore, diffusion length $l_d (=2D/v)$ is in inverse proportion to the 9th power of supercooling Δ. Table 1 shows the relation of supercooling Δ, diffusion length l_d and the number of doublon that survives finally N_S. Because the diffusion length l_d is an amount that expresses a length of distance influenced by diffusion field, as the l_d decreases, the number of surviving doublon increases. When L_x=320, N_s decreases compared with L_x=480 at same Δ. The reason is that as L_x decreases, the distance of doublons lowers and it is the same effect as that of increasing the diffusion length.

Table 1 The relation of supercooling Δ, diffusion length l_d and number of doublon that survives finally N_S.

| Δ	l_d	$N_S(L_x$=480)	$N_S(L_x$=320)
0.70	39.6	1	1
0.71	34.9	1	1
0.72	31.0	1	1
0.73	27.4	1	1
0.74	24.2	1	1
0.75	21.6	1	1
0.76	19.2	2	1
0.77	17.0	2	1
0.78	15.2	2	1
0.79	13.4	2	2
0.80	11.7	3	3
0.81	9.73	3	3

When Δ exceeds threshold (Δ =0.8), the property of the group of doublon changes. Figures 3 (a), (b), (c) and (d) are, respectively, first, second, third and fourth simulations at Δ =0.82 and L_x=320. Figure 3(a) shows that the grooves of doublons are initially buried because of strong Δ [12]. However, each doublon makes new groove between other doublons. As they grow, all fingers approach each other and some grooves of their fingers begin to oscillate. Even though the grooves looks oscillating irregularly, whole envelope keeps stable form. This form looks like DBM and to investigate this pattern is future work.
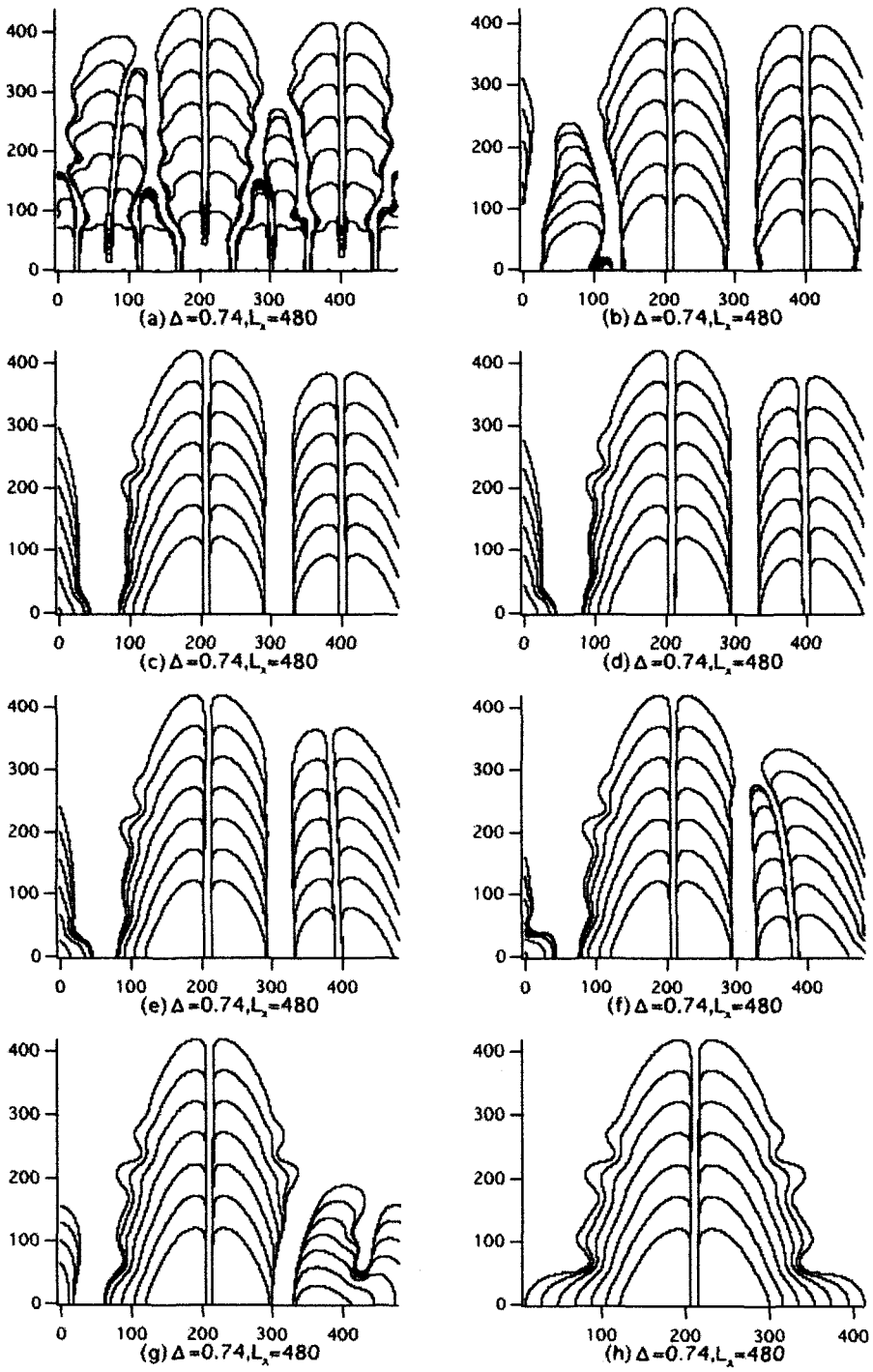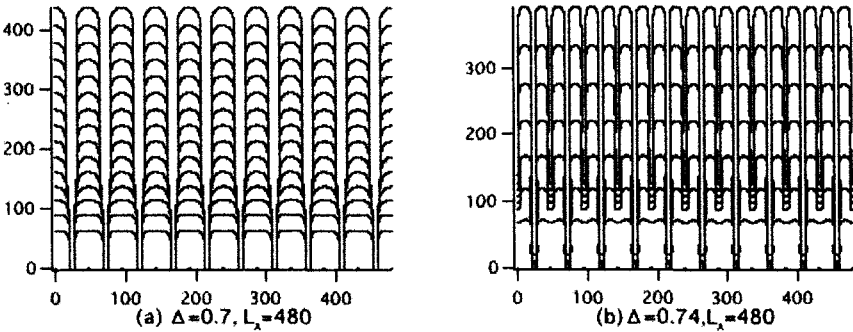
Figs 1: Group of doublon patterns at $\Delta = 0.74$ and $L_x = 480$
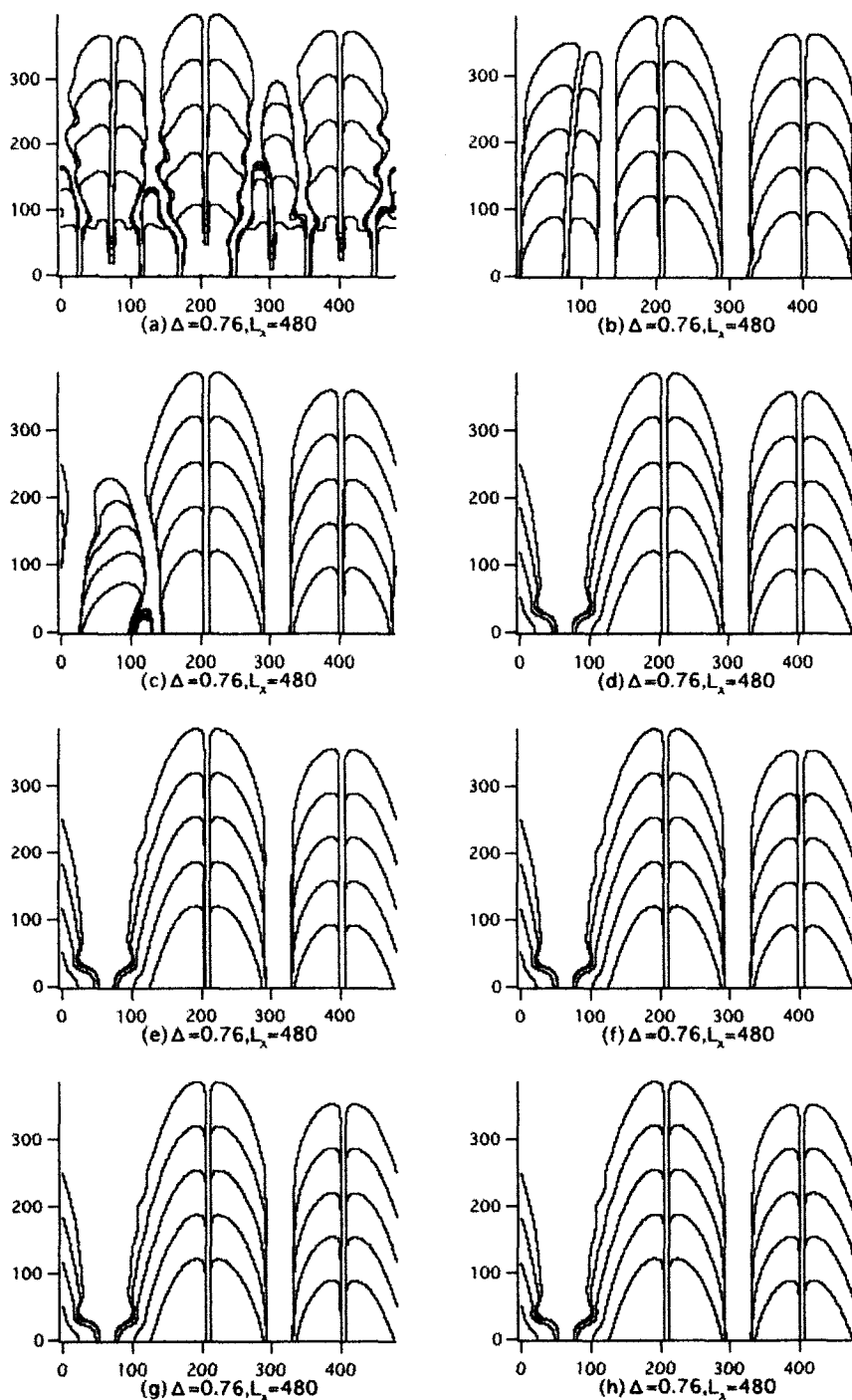
Figs 2: Group of doublon patterns at $\Delta = 0.76$ and $L_x = 480$
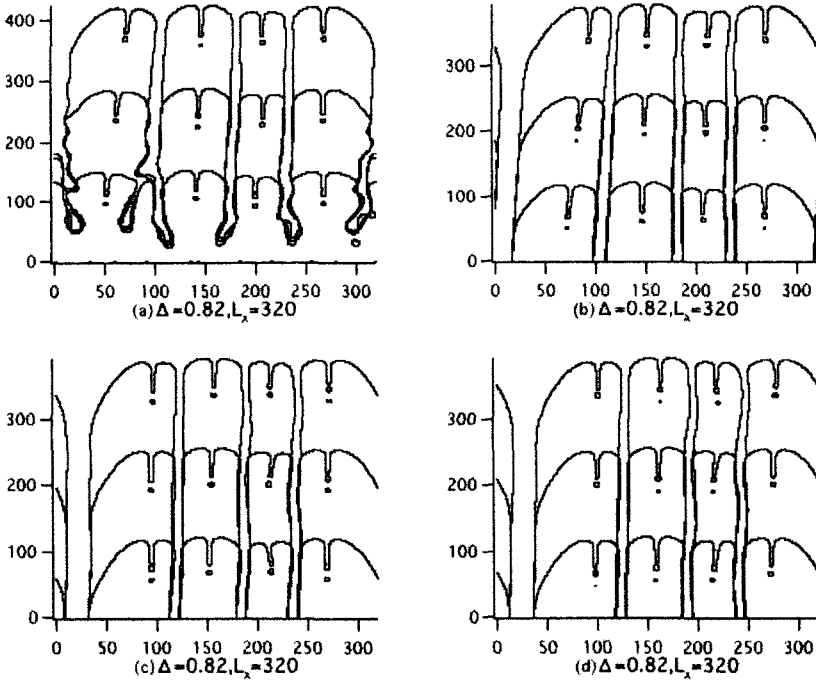
Figs. 3: Growing of DBM-like pattern at $\Delta=0.82$, $L_x=320$.

Figs. 4: Patterns grown from 10 same size crystals that is at regular interval arrangement

The appearance of these competition of doublons needs initial perturbation. In Figs. 4, we show some patterns grown from 10 same size crystals that is at regular interval arrangement. In Fig. 4(a) at $\Delta=0.7$ and $L_x=480$, all the fingers go straight constantly. Figure 4(b) at $\Delta=0.74$ and $L_x=480$ is also grown from regular interval initial crystal arrangement. Tip splitting is occurred in all the fingers. And all the fingers are growing evenly just like the growth in narrow channel.

5. Summary

In this paper, we observe the doublons growing under the effect of other doublons. In isotropic cases, a doublon pattern cannot grow up stably when another doublon is growing up ahead. As time passes, the number of growing doublon decreases, affected by competition of other doublons. It has been understood that the number of doublons that survives to the last minute depends on the diffusion length.

When Δ exceeds some value, the groove of doublon vanishes and makes new groove between other fingers. This form has stable envelop and whole shape looks like DBM.

Appendix
A1. Discretization

The temperature and the order parameter of the phase-field model change in space continuously. However, for the numerical simulation, we have to discretize diffusion field of the temperature and order parameter. The growth of doublons depends on the anisotropy very sensitively.

The simplest discretization of $\nabla^2 u$ is

$$\{u_n(i+1,j) + u_n(i-1,j) + u_n(i,j+1)$$
$$+u_n(i,j-1) - 4u_n(i,j)\}/\Delta x^2. \tag{A1}$$

The term n is the time step number and (i,j) denotes the lattice point. However, this scheme of the spatial discretization is not sufficient, because numerical anisotropy cannot be neglected. In the case of simulation including surface tension anisotropy and kinetic anisotropy, that numerical anisotropy is very small compared with these anisotropies and can be neglected. In this paper, because we would like to simulate more isotropic case, we use new discretization expressed as

$$\nabla^2 u = D\{u_n(i+1,j+1) + u_n(i-1,j-1)$$
$$+ u_n(i-1,j+1) + u_n(i+1,j-1)$$
$$- 4u_n(i,j)\}/\Delta x^2. \tag{A2}$$

$$\nabla^2 u \sim a(A1) + b(A2). \tag{A3}$$

By tuning the parameters as $a=2/3$ and $b=1/6$, the numerical scheme becomes more isotropic than Eq.(A1). In this paper, we use the method of discretization to investigate the doublon in more isotropic case.

Reference

1) S. K. Chan, H. H. Reimer and M. Kahlweit, J. Cryst. Growth, 32(1976), 303.
2) S. C. Huang and M. E. Glicksman, Acta Metal, 29(1981) 714.
3) A. Dougherty, P. D. Kaplan and J.P. Gollub, Phys. Rev. Lett., 58(1987) 1652.
4) S. Akamatsu, G. Faivre and T. Ihle, Phys. Rev. E, 51(1995) 4751.
5) W. Losert, D. A. Stillman, H. Z. Cummins, P. Kopczn'ski, W. –J. Rappel and A. Karma, Phys. Rev. E, 58(1998) 7492.
6) N. Samid-Merzel, S. G. Lipson and D. S. Tannhauser, Physica A, 257(1998) 413.
7) H. M. Singer and J. H. Bilgram, Phys. Rev. E, 70(2004) 031601.
8) R. Kobayashi, Physica D, 63 (1993) 410.
9) A. A. Wheeler, B. T. Murray and R. J. Schaefer, Physica D, 66(1993) 243.
10) B. T. Murray A. A. Wheeler and M. E. Glicksman, J. Crystal Growth,154(1995) 386.
11) A. Karma and W. J. Rapell, Phys. Rev. E, 53(1996) 3017.
12) S. Tokunaga and H. Sakaguchi, Phys. Rev. E, 70(2004) 011607.
13) S. Tokunaga and H. Sakaguchi, in: M.M. Novak (Ed.), Thinking in patterns, World scientific, Singapore, 2004,pp. 301.
14) S. Tokunaga and H. Sakaguchi, J. Crystal Growth, 275(2005) e197.
15) M. B. Amar and E. Brener, Phys. Rev. Lett., 75(1995), 561.

This page is intentionally left blank

STUDY OF THERMAL FIELD IN COMPOSITE MATERIALS

PAVLA STEFKOVA, OLDRICH ZMESKAL, RADEK CAPOUSEK

Institute of Physical and Applied Chemistry, Faculty of Chemistry, Brno University of Technology, Purkynova 118, 612 00 Brno, Czech Republic

A new method of composite materials thermal parameters determination is described in this contribution. The transient pulse method is used for specific heat, thermal diffusivity and thermal conductivity determination. The evaluation is performed with the help of mathematical apparatus used study properties of fractal structures. The method presented is more general then the classical transient method commonly used for description of homogenous materials.

1 Introduction

The Fourier laws usually describe *the heat conduction* in solids; their stationary forms can be written with the help of equation [1, 2, 3]

$$q_0(r) = \operatorname{div} q = \operatorname{div}(-\lambda \operatorname{grad} T) = -\lambda \Delta T , \qquad (1)$$

where $q_0(r)$ is the specific yield of the heat source, $q(r)$ is the density of heat flow rate, $T(r)$ is the temperature and λ is the constant thermal conductivity and equation

$$\lambda = c_p \rho a = C_p a , \qquad (2)$$

where c_p is the specific heat capacity at constant pressure, C_p is the density of heat capacity, ρ is the mass density and a is the coefficient of thermal diffusivity of the body.

The main aim of this contribution is to formulate a uniform mathematical system, which can describe the heat transfer independently of the properties of the matter. Fractal dimension and fractal measure will be the only parameters distinguishing the differences.

Let's choose the density of the heat capacity Cp as a basic quantity. It was found [4] that in fractal environment the temperature dependency of the heat capacity Cp density can be described as

$$C_p(r) = c_p(r)\rho(r) = \frac{\lambda}{a} = -\frac{k^2}{\hbar c} \Delta T , \qquad (3)$$

where $k = 1.380658 \times 10^{-23} \, \mathrm{J K^{-1}}$ is the quantum of heat capacity (Boltzmann constant), $\hbar = 1.05457266 \times 10^{-34} \, \mathrm{J s}$ is the modified Planck's constant and $c = 2.997925 \times 10^8 \, \mathrm{m s^{-1}}$ is the speed of light in vacuum. The density of fractal physical quantity $C_p(r)$ in E - dimensional Euclidean space E_n $(E = n)$ (the density of heat capacity) was defined [5, 6] as

$$C_p(r) = k \, n(r) = k \, K \, r^{D-E} , \qquad (4)$$

where $n(r)$ is the coverage of space (distribution of particles concentration), r is the radius of elementary quantity, K the fractal measure (number of elements in volume with

$r = 1$) and D is the fractal dimension. It is apparent already from this equation that in equally filled space ($D = E$) the heat density is constant, while for the point source of the heat ($D = 0$, $E = 3$) the heat density decreases with the distance from the source with third power. From (4) we can calculate the heat capacity $C_s(r)$ using equation

$$C_s(r) = k N(r) = \int_{V^*} C_p(r) \, dV^* = k K \frac{E r^D}{D}, \tag{5}$$

where $N(r)$ is the number of particles in volume $V^* = r^E$ and $dV^* = d(r^E)$ is the elementary volume of E-dimensional space. For radial temperature field we can write the dependence of temperature on radius using (4)

$$T_r(r) = -\frac{\hbar c}{k} \frac{K r^{D-E+2}}{D(D-E+2)}. \tag{6}$$

If we suppose that the heat permeates through the surrounding by the constant diffusivity in time $r^2 = 4a_0 t - r_T^2$, where $r_T^2 = x^2 + y^2 + z^2$ is a radius of fractal space and a_0 is the maximum value of diffusivity (for $D = E$). We get from (6) [8, 9]

$$T_r = -\frac{\hbar c}{k} \frac{K(4a_0 t)^{(D-E+2)/2}}{D(D-E+2)} \left(1 - \frac{r_T^2}{4a_0 t}\right)^{(D-E+2)/2}. \tag{7}$$

If the heat diffuses by the significantly lesser speed ($r_T^2 \ll a_0 t$, small distances or long times), then the terms in parenthesis can be observed as significant in the expansion of exponential function ($1 - x \approx e^{-x}$) and we can write

$$T_r = -\frac{\hbar c}{k} \frac{K(4a_0 t)^{(D-E+2)/2}}{D(D-E+2)} \exp\left(-\frac{D-E+2}{2} \frac{r_T^2}{4a_0 t}\right). \tag{8}$$

If we substitute for the thermal diffusivity (2) $a = 2 a_0 /(D-E+2)$, where a is the coefficient of thermal diffusivity of the body with fractal dimension D, we obtain

$$T_r = -\frac{\hbar c}{k} \frac{2atK}{D} \left[2at(D-E+2)\right]^{(D-E)/2} \exp\left(-\frac{r_T^2}{4at}\right). \tag{9}$$

For the total heat transferred to the body from the heat source

$$Q = -\frac{2\hbar c}{k} \frac{Kt}{D[(D-E+2)/2\pi]^{(D-E)/2}} \tag{10}$$

we can write

$$T_r = \frac{Q}{c_p \rho (4\pi a t)^{(E-D)/2}} \exp\left(-\frac{r_T^2}{4at}\right). \tag{11}$$

The relation (11) is applicable [1, 2, 3] for fractal dimensions $D = 0, 1, 2$ and topological dimension $E = 3$, see Figure 1.
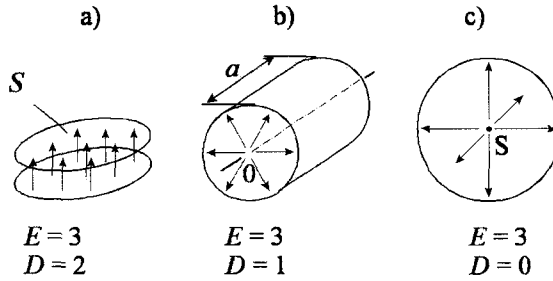
Figure 1. Heat flow geometry for a) plane-parallel, b) cylindrical and c) spherical coordinates Euclidean space.

2 Model calculations

The maximum position can be determined by the derivation of (11) with the time

$$\frac{\partial \log T}{\partial \log t} = \left(\frac{D - E}{2} + \frac{h^2}{4at} \right) = 0 . \tag{12}$$

From this equation the thermal diffusivity at the maximal time can be determined

$$a = \frac{h^2}{2t_m f_a} = \frac{h^2}{2(E - D)t_m} , \tag{13}$$

where f_a is a coefficient that characterizes the deformation of the thermal field [2]. This coefficient is equal to one for the ideal plane source ($E = 3, D = 2$). The maximum temperature of the response for the Dirac thermal pulse is obtained by introducing of the thermal diffusivity (13) in (11)

$$T_m = \frac{Q}{c_p \rho} \exp\left(\frac{D - E}{2} \right) \cdot \left(\frac{E - D}{2\pi h^2} \right)^{(E-D)/2} . \tag{14}$$

From the ratio of equations (14) and (11) and with the use of (13)

$$\frac{T_m}{T} = \left[\frac{t}{t_m} \exp\left(\frac{t_m}{t} - 1 \right) \right]^{\frac{E-D}{2}} , \tag{15}$$

it is possible to define the coefficient f_a (fractal dimension D respectively) for every point of the experimental dependence

$$f_a = E - D = \frac{2 \ln(T_m / T)}{\ln(t/t_m) + (t_m/t - 1)} . \tag{16}$$

The value of the coefficient f_a could be also affected by the geometry of sample [2] or by the finite pulse width [4].

The thermal capacity can be derived from Eq. (2):

$$c_p = \frac{Q}{\rho T_m h_T} \frac{f_c}{\sqrt{2\pi \exp(1)}} = \frac{Q}{\rho T_m h^{E-D}} \left(\frac{E-D}{2\pi \exp(1)} \right)^{(E-D)/2} \tag{17}$$

as well as the thermal conductivity of the studied fractal structure:

$$\lambda = c_p \rho\, a = \frac{Q}{2(E-D)T_m t_m h^{E-D-2}} \left(\frac{E-D}{2\pi \exp(1)} \right)^{(E-D)/2}, \tag{18}$$

where $\exp(1)$ is Euler number and f_a, f_c are the coefficients that characterize the deformation of the thermal field [2].
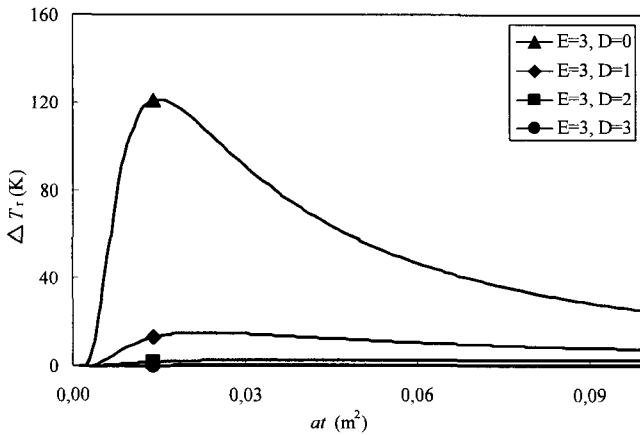
Figure 2. Time dependencies of the model temperature responses for the Dirac thermal pulse (for the heat flow geometry from Figure 1 calculated by Eq. (11)

Figure 3 represents time-temperature dependencies (c.f. (11)) calculated for spherical ($D = 0$), cylindrical ($D = 1$), planar ($D = 2$), and "cubic" ($D = 3$) geometry of the heat source (see Figure 1). From these dependencies the fractal dimension was reconstructed with help of Eq. (16). The results are in the Figure 3. The differences between input fractal dimensions and the calculated ones are in the whole time interval negligible. It is caused by the fact that the term (16) can't be numerically calculated for $t \to t_m$. When the value f_a (fractal dimension D, respectively) is known, it is feasible to determine the parameters of the studied thermal system with the aid of the (13) – (18) equations.
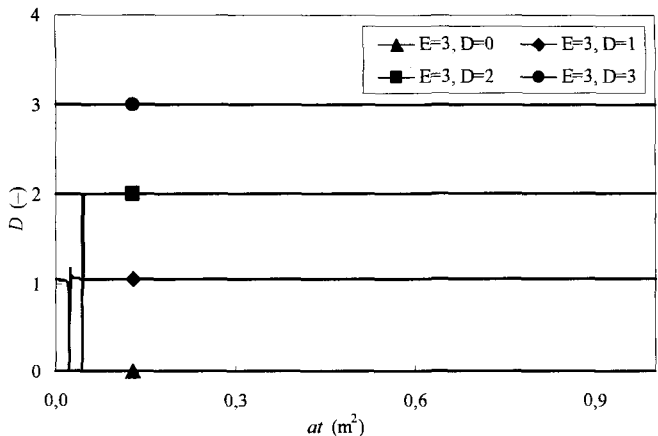
Figure 3. Time dependencies of the reconstructed fractal dimensions (16) calculated from temperature responses to the Dirac thermal pulse (for the heat flow geometry from Figure 1 calculated by Eq. (11)

3 Experimental

Measurements of the responses to the pulse heat were performed using the Thermophysical Transient Tester 1.02 developed at the Institute of Physics, Slovak Academy of Science. The measured sample of composite thermal insulating material was round with the diameter $R = 0.03$ m. Its density was $\rho = 247.885$ kg m^{-3} for its thickness $h = 8.2$ mm.
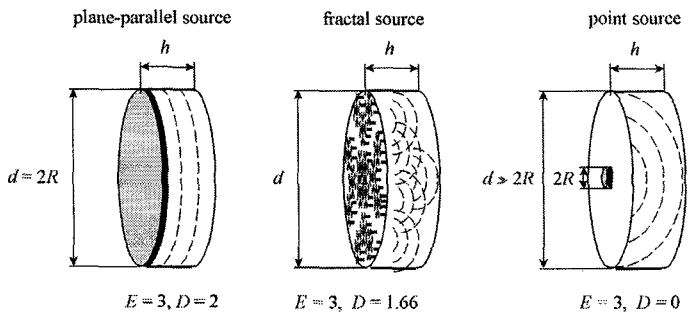
Figure 4. Current flow geometry: a) plane-parallel, b) fractal, c) point (for different ratio of length contact respectively) source.

In Figure 4 three possible configurations of the experimental arrangements are shown. In Figure 4a the diameter of the specimen is equal to the diameter of the heat source, in Figure 4c, diameter of a heat source is far smaller than specimen's diameter. Figure 4b shows the real situation, when the heat is delivered irregularly (either from the source of finite size (capacity) or from a source with specific composition of heat sources – e.g. hot-disc).

4   Results

The typical time responses of temperature for the rectangle (Dirac) pulse of different input power with width $\Delta t = 2$ ms are presented in Figure 5. The corresponding heat energy was changed from 191 J m^{-2} to 1944 J m^{-2}. From this figure it is evident that the position of the dependence maximum t_m changes slowly, but the maximum value of temperature T_m is increased.
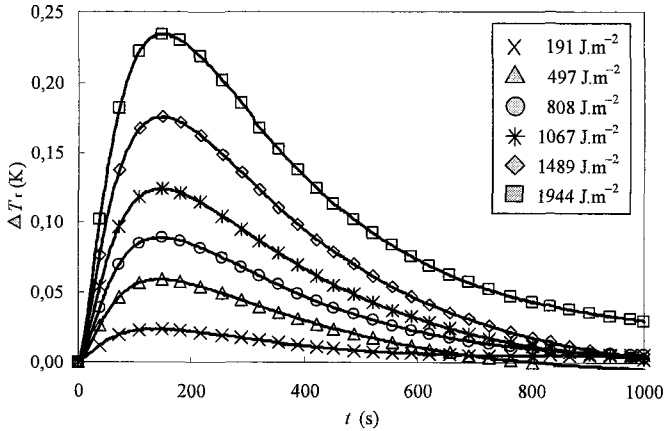
Figure 5. Thermal responses of the sample measured by the pulse transient method for different pulse power
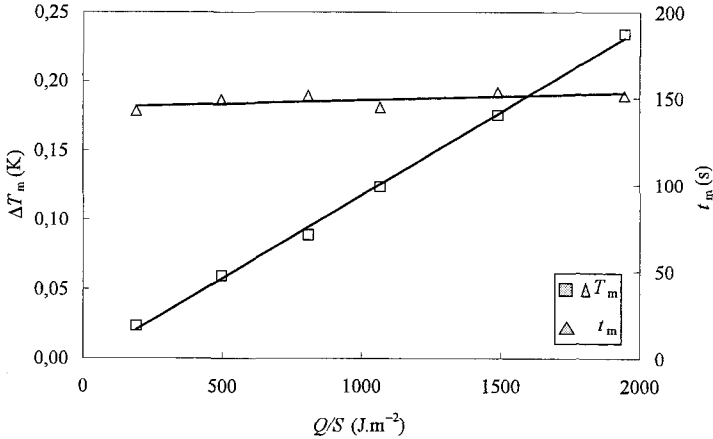
Figure 6. The dependences of the time of characteristics' maximums and maximums temperature values on the heat energy.

The detailed dependences of the time corresponding to the characteristics' maxima and maximal temperature values of the heat energy of the source are in Figure 6. It is evident that the positions of peaks increased slowly, such that the diffusivity is approximately constant, see (13). The dependencies of fractal dimensions were calculated with the help of Eq. (16) (analogically to the fractal dimensions calculated for

models from Figure 2 in Figure 3). They are presented in Figure 7. It is evident, that these dependencies are very different for model characteristics (Figure 3). It is caused probably by the losses of heat during its propagation through the composite material.

From the Figure 7 it is evident that for very short times there is a value of the fractal dimension $D \approx 3$ and therefore, the plane heat source is formed. The value of the fractal dimension decreases with increasing time, since the heat disperses into the space.
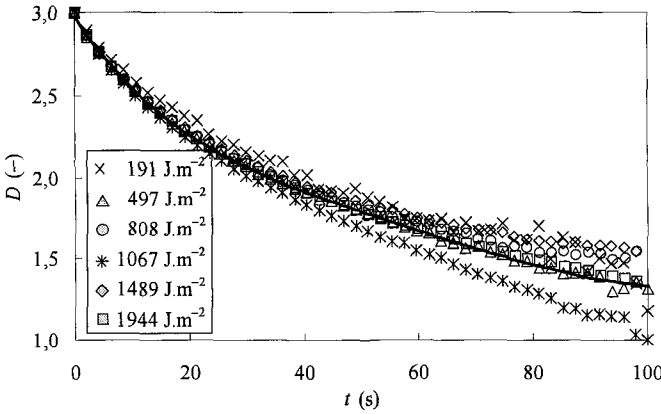
Figure 7. Fractal dimension of the heat distribution in the specimen determined from increased part of characteristics.

The final parameters: thermal diffusivity (13), specific heat capacity (17) and thermal conductivity (18) for planar heat source ($D = 2$) are presented in Table 1. It is evident that the results are not dependent on the heat energy of source. The statistical error is less than 4 % for all three cases. However, the situation is more complex because of the heat losses and also because the measured material was inhomogeneous.

Table 1: Thermal parameters of studied insulating composite material

Q/S (J m^{-2})	t_m (s)	ΔT_m (K)	a (m^2 s^{-1})	c_p (J kg^{-1} K^{-1})	λ (W m^{-1} K^{-1})
191.13	142.75	0.024	2.36×10^{-7}	957.26	0.056
496.99	149.30	0.060	2.25×10^{-7}	990.99	0.055
808.20	151.40	0.089	2.22×10^{-7}	1078.32	0.059
1066.81	144.86	0.124	2.32×10^{-7}	1025.34	0.059
1489.58	153.53	0.176	2.19×10^{-7}	1009.60	0.055
1944.02	151.33	0.234	2.22×10^{-7}	988.72	0.054
			$(2.26 \pm 0.06) \times 10^{-7}$	(1008.37 ± 37.60)	(0.056 ± 0.002)

Conclusion

The results of thermal responses to the Dirac pulse of supplied heat are discussed. Simplified heat conductivity model was used [7] to interpret the experimental outcomes. The model is based on published [1] expectations. The results show the image of heat

distribution in the specimen for various heat powers from the source. These evaluations could be used for more accurate determination of the thermal parameters of materials. The differences between model calculations and experimental results are evidently caused by the inhomogeneities of the composite material studied and by the heat loss during the measurement. These discrepancies will be studied in more detail in our further work.

Acknowledgments

References

1. H. S. Carslaw, J. C. Jaeger, *Conduction of Heat in Solids*. Clarendon Press London (1959)
2. J. Krempaský, *Measurement of Thermophysical Quantities*. VEDA Bratislava (1969)
3. Ľ. Kubičár, *Pulse Method of Measuring Basic Thermophysical Parameters*. VEDA Bratislava and Elsevier Nederland (1990)
4. O. Zmeskal, M. Buchnicek, M. Vala, *Chaos, Solitons & Fractals* 25, 941 (2005)
5. O. Zmeskal, M. Nezadal, M. Buchnicek, *Chaos, Solitons & Fractals* 17, 113 (2003)
6. O. Zmeskal, M. Nezadal, M. Buchnicek, *Chaos, Solitons & Fractals* 19, 1013 (2004)
7. V. Boháč, Ľ. Kubičár, V. Vretenár, *TEMPMEKO, 9th International Symposium on Temperature and Thermal Measurements in Industry and Science*, Dubrovnik (2004)
8. O. Zmeskal, M. Buchnicek, M. Nezadal, P. Stefková, R. Capousek, *Thermophysics 2003*, Kočovce (2003).
9. O. Zmeskal, P. Stefková, *Thermophysics 2004*, Kočovce (2004).

SIMULATION OF GEOCHEMICAL BANDING IN ACIDIZATION-PRECIPITATION EXPERIMENTS IN-SITU*

MAYSAM MSHARRAFIEH, MAZEN AL-GHOUL AND RABIH SULTAN

Department of Chemistry
American University of Beirut
Beirut, Lebanon
E-mail: rsultan@aub.edu.lb
Fax: 961-1-365217

This paper explores the possible similarities between the well-known Liesegang banding phenomenon in precipitate systems, and the stripe formation observed in a large number of rocks. We present new experiments performed in-situ (real rock systems), to simulate the band formation through the acidization of a ferruginous limestone rock, causing dissolution and precipitation reactions. The results are analyzed by microscopy, AA and XRD techniques. The correlation between the Liesegang gel experiment and the processes taking place inside the rock medium is established. A theoretical model to support our experiments is also presented.

1. Introduction

Periodic precipitation, known as Liesegang banding[1,2], occurs when two co-precipitate ions interdiffuse in a gel medium. This occurs through the coupling of the precipitation reaction with diffusion. As a result, spatial oscillations through periodic precipitation are displayed in the form of bands in 1D, or in the form of rings in 2D. Similarities with the Liesegang phenomenon mainly include the banded patterns found in a vast variety of rocks (geochemical self-organization). Banded rock formations arise from the cyclic precipitation, as mineral-rich water infiltrates a porous rock and reacts to form an insoluble product[3].

Examples of rock types whose textures are attributed to such processes include: banded iron oxide minerals, zebra spar[4], wood-grained chert[3], and banded agates and geodes[5,6]. Hematite rings in sandstones are an example of rhythmic banding similar to a Liesegang type experiment. Hematite (cherry red Fe_2O_3) is a characteristic mineral of Liesegang rocks occurring as ring-shaped aggregates several tens of microns in diameter[7]. Another commonly cited example of Liesegang bands in rocks is the striped precipitation of iron oxide minerals upstream of an oxidative front, lagging behind a dissolution front within a pyrite (FeS_2) rock bed. The scenario

leads to the banded deposition of goethite ($Fe^{3+}O(OH)$) behind the dissolution front[8]. A mechanism for this type of band formation was set forth[8], based on a bifurcation the Ostwald supersaturation-nucleation-depletion cycle[9].

Agates represent one of the classic types of rocks whose self-organized structure was attributed to the Liesegang mechanism of banded precipitation. They consist of microcrystalline SiO_2 tight aggregates that display a rich colored structure due to the impurities present in the SiO_2 matrix. The mechanism yielding ring-shaped domains is attributed to the composition of agate which varies greatly. It consists of microcrystalline silica such as chalcedony, carnelian, amethyst, jasper, opal, and flint, often occurring as layers in agate which are deposited from aqueous solutions[10]. The experimental simulation of the moss agate formation by precipitating iron (II) silicate solution in water glass showed that reaction-diffusion processes within the magma may cause the growth of ring-shaped domains that yield agate-type rock upon cooling of the melt[11].

Whereas theoretical modeling attempts to simulate geochemical self-organization exist in the literature[8,12,13], experimental simulations are scarce and are limited to Liesegang-type experiments in gels, not in real rocks. The Liesegang gel method was used to model gallstone formation *in vitro*, approximating the conditions present in bile salt systems[14,15]. In this study, we propose to perform acidization-precipitation experiments *in-situ*, using porous ferruginous limestone rock samples. We then present a reaction-transport model to initiate a theoretical study that will serve as a basis for the mechanistic unraveling of the observed phenomena.

Rock acidization studies were conducted in dolomite[16], and sandstone cores[17], but no attention to pattern formation was given. The proposed experiment consists of allowing an acid to slowly infiltrate through a rock medium. The acid causes the dissolution of parts of the latter, and the subsequent precipitation of another mineral behind the dissolution front. We anticipate the precipitation will occur in the form of bands. Consider a rock sample of limestone rich in calcite mineral ($CaCO_3$). A flow of acidic water causes the dissolution of the rock according to the following equation:

$$CaCO_3(s) + 2H^+(aq) \longrightarrow Ca^{2+}(aq) + CO_2(g) + H_2O(l) \tag{1}$$
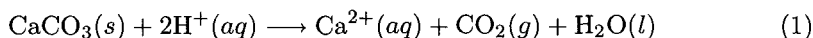
The presence of anions such as, say, SO_4^{2-} in the infiltration flow can precipitate anhydrite ($CaSO_4$) or gypsum ($CaSO_4 \cdot 2H_2O$), according to the reaction:

$$Ca^{2+}(aq) + SO_4^{2-}(aq) \longrightarrow CaSO_4(s) \tag{2}$$

2. The Experiment

2.1. *Setup*

A red brown, porous ferruginous limestone rock was used in the experiment. A 4.0 M H_2SO_4 solution was allowed to infiltrate into the rock by means of an infusion pump (Thermo-Orion M361 Multi-rate Infusion Pump). The pump syringe was filled with 50 mL of H_2SO_4, and the pump was set to deliver the acid at the rate of 1.0 mL per hour, through a Tygon tube fitted into a tiny orifice (3 mm diameter, and 1 cm depth) carefully drilled in the rock. With this injection setup, a well-controlled flow could be attained. The pump syringe was repeatedly refilled, as soon as the acid was consumed.

2.2. *Results and discussion*

Texture and color alterations within the ferruginous limestone rock (by comparison before and after acid injection) gives a preliminary idea on the chemical processes that have taken place. A photomicrograph illustrating such modifications is depicted in Fig. 1. Frame **a** shows the bare rock before the reaction, while the acid-invaded medium is seen in frame **b.**
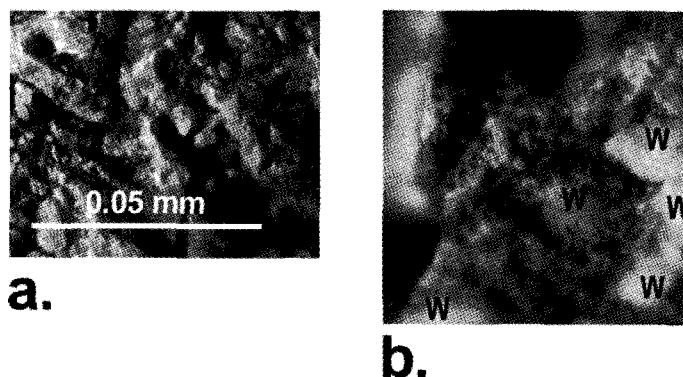
Figure 1. Rock texture before (a) and after (b) acid injection. The latter displays the white crystals of $CaSO_4$ (labeled W) embedded within the rock. The two photomicrographs are of the same magnification (10×). The scale is indicated on the picture.

The white crystals, labeled "W", embedded within the rock can be distinctly remarked after the dissolution and precipitation reactions (frame **b**), and consist of the salt $CaSO_4$, precipitated by reaction 2. The rock has the appearance shown in Fig. 2 after 161 days of steady acid injection (4.0 M H_2SO_4).

The white precipitate is first concentrated in the central region, spanning a circular zone surrounding the injection center. To do a preliminary test for the presence of any remaining calcium carbonate in this deposited white zone, a drop
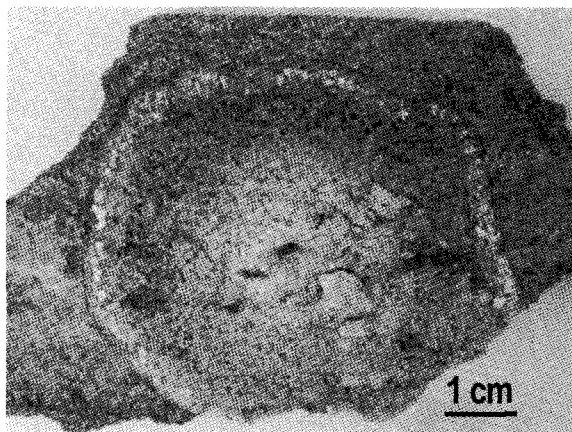
Figure 2. $CaSO_4$ deposition in the limetsone rock sample, infiltrated by H_2SO_4. The central region is a dense precipitate zone. Then, beyond a brown depleted zone, a distinct belt of $CaSO_4$ suggests the formation of a first Liesegang ring. The photo is taken on day 161. A 1 cm scale is indicated on the picture.

of concentrated acid (HCl) is added. No effervescence is produced indicating the depletion of the region from carbonates. This will later be confirmed by X-ray. Effervescence is due to CO_2 produced when the acid reacts with the carbonate ions (CO_3^{2-}). Beyond this region, another region appears, delineated by the boundary of the infiltration front. Essentially, no white precipitate (similar to that in the center) is observed in this region. We will refer to this region as the "inter-band zone". No CO_2 effervescence occurs there upon HCl addition, also suggesting a depletion from the presence of carbonate salts (consumed by dissolution). Following the inter-band zone, comes a distinct ring of white precipitate (with traces of yellow). This belt of mainly white precipitate is believed to mimic a first Liesegang band, formed after the depleted inter-band zone. A photomicrograph of the rock around that ring (depicted in Fig. 3) clearly reveals a difference in texture between it and both the bare rock and the acid-reacted zone. This sharp contrast between the ring domain and the medium surrounding it favors the hypothesis of an individualized Liesegang band. The picture was taken with a Nikon (COOLPIX 4500) digital camera, observing through the lens of a fluorescence microscope (Zeiss AXIOVERT 200). The front edge is indicated by the red arrows on the Figure (boundary between the dark and light brown regions). So we see that the acidity front has reached a location slightly after the first ring.

We now attempt to identify the salts formed in the different zones of the rock (notably the central region and the ring domain). Simple tests were first conducted. Samples of white precipitate from both the central region and the peripheral ring were treated with HCl. No fizzing indicative of chemical reaction of acid with carbonate was noticed for the white precipitate in the central region. A small sample
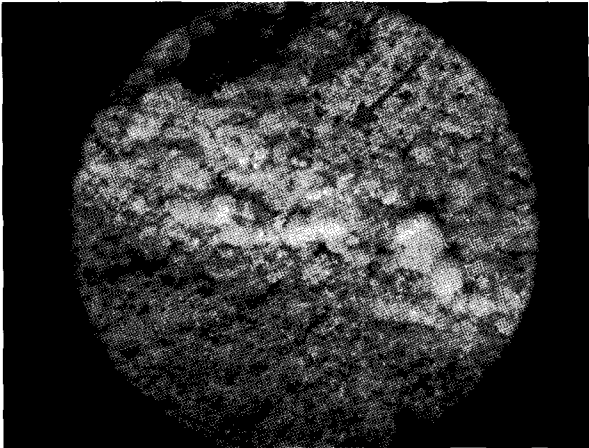
Figure 3. Photomicrograph showing the texture contrast between the bare rock (upper portion) and the acid-invaded zone (lower portion), demarcated by the white and slightly yellow crystals (at the center). The picture is obtained at a 4× optical zoom magnification. The diameter of the displayed rock portion is approximately 1.5 cm. The arrows indicate the position of the front, moving upwards.

Table 1. Analysis of the ring domain for Ca and Fe obtained by flame atomic absorption analysis. The second and third columns indicate the concentrations in mg/ L (ppm) of Fe and Ca. These concentrations are converted to masses, and the mass ratio of calcium to iron is finally calculated (last column).

Sample ID	Fe (ppm)	Ca (ppm)	m_{Fe} (mg)	m_{Ca} (mg)	m_{Ca}/m_{Fe}
Center	3.528	36.19	0.0353	0.3620	10.3
Ring	9.629	8.947	0.09626	0.08947	1.08

of white salt from the ring dissolved in HCl, yielding a yellow solution, attributed to the presence of ferric ions originating from the rock medium (ferruginous limestone). The presence of Fe^{3+} was confirmed by testing with KSCN solution. Upon addition of a few drops of SCN^-, the solution turned blood-red indicating the formation of the $FeSCN^{2+}$ complex. Flame atomic absorption analysis of the precipitate (done on a Thermo-elemental Solaar atomic absorption spectrophotometer), in both the ring and the central region (after dissolution in HCl) was then undertaken to determine the elemental mass ratio of calcium to iron. The results are recorded in Table 1.

As can be inferred from the above results, the central region is much richer in calcium ions confirming that the obtained precipitate is overwhelmingly calcium sulfate. In the ring domain, it is established that both iron and calcium are present in almost a 1:1 ratio. A sample of white salt from the ring zone (traces) dissolved in water; but its fairly high calcium content still points toward the possible presence of calcium sulfate in the ring. For a rigorous characterization of the different rock regions (before and after acid injection), we now carry out an X-ray crystallo-

graphic analysis. This aids in the understanding of the chemical processes that took place inside the rock due to the H_2SO_4 infiltration, and verifying the mechanism of Liesegang ring formation.

2.3. X-Ray analysis

Powder X-ray diffraction data is collected using the reflection geometry. Samples of the white salt from the central region (near the injection site), and the ring were analyzed using X-ray crystallographic techniques. A sample from the bare rock was also studied. The X-ray diffractometer (Bruker D8 Discover) was run at 40 kV and 40 mA. The steps are in increments of 0.01 degrees, and counts were collected at 0.5 seconds per step. The vertical axis records the X-ray intensity, while the horizontal axis records the scanned 2θ angles, in degrees. Low angles (large d-spacings) lie to the left. The peak positions and their intensities provide important information about the structure of the material. The plots are shown in Figs. 4, 5 and 6.

We search for calcite ($CaCO_3$), gypsum ($CaSO_4.2H_2O$), and anhydrite ($CaSO_4$) in each of the sampled areas. The X-ray results for the bare rock sample show that it is rich in calcite ($CaCO_3$) with only very few traces of gypsum. This is inferred from Fig. 4 and Table 2.
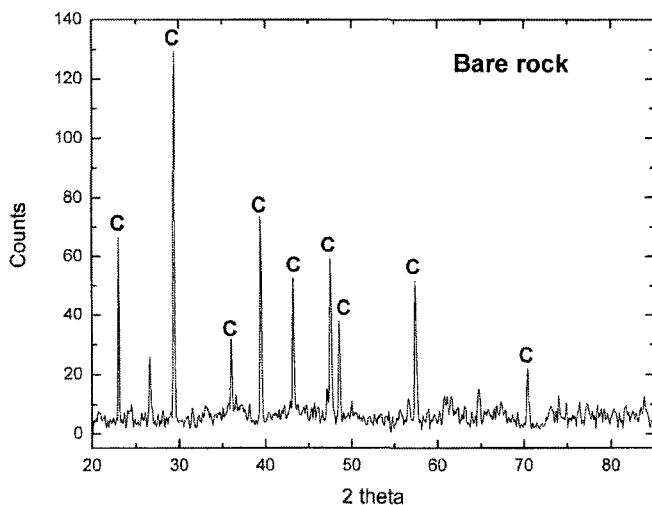
Figure 4. X-ray diffraction spectrum of the bare rock. The rock is overwhelmingly composed of the calcite ($CaCO_3$) mineral (C peaks).

The analysis of the white precipitate in the central region reveals that it is predominantly gypsum, and to a lesser extent, anhydrite (see Fig. 5). Traces of monohydrocalcite ($CaCO_3 \cdot H_2O$) were also identified (two low intensity peaks at
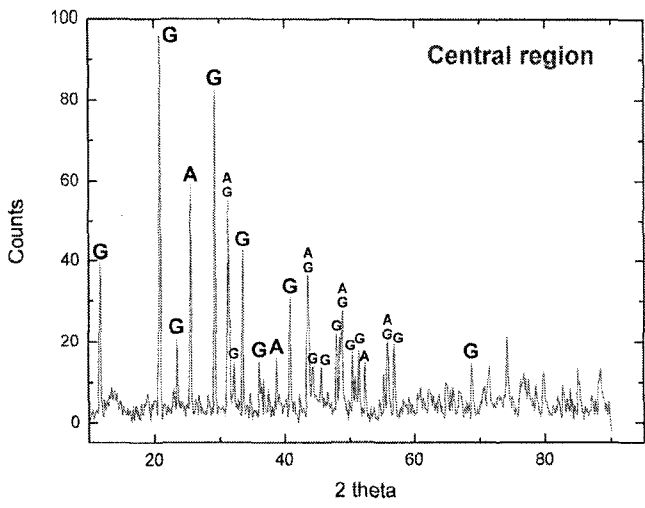
Figure 5. X-ray diffraction spectrum of the central region after the dissolution-precipitation reactions. The extensive formation of $CaSO_4$ as anhydrite (A) and gypsum (G) is evident.

Table 2. Relative abundance of $CaSO_4$ and $CaCO_3$ in the central region, the remote ring (1st Liesegang band) and the bare rock, determined by powder X-ray diffraction analysis. The presence of $CaSO_4$ in the ring is confirmed.

	Bare Rock	Central Region	Ring
Total counts (gypsum + anhydrite)	36.4	742.8	279.3
Total counts of calcite	569	35.8	114.9
Ratio $CaSO_4/CaCO_3$	0.0640	20.7	2.43

$2\theta = 36.2°$ and $51.5°$). Note that the calcite, anhydrite and gypsum peaks (C, A and G respectively) all lie in the region between $2\theta = 20°$ and $60°$. Table 2 shows that in the central region, $CaSO_4$ (as gypsum and anhydrite) is overwhelmingly dominant over $CaCO_3$, which has essentially dissolved by virtue of reaction 1.

Finally, the presence of $CaSO_4$ in the ring is confirmed (see the gypsum peaks in Fig. 6). Table 2 shows that gypsum is dominant over calcite. This realization is the most important result in the study, as it reveals to a great extent the viability of a Liesegang-type scenario in certain pattern forming processes in rocks.
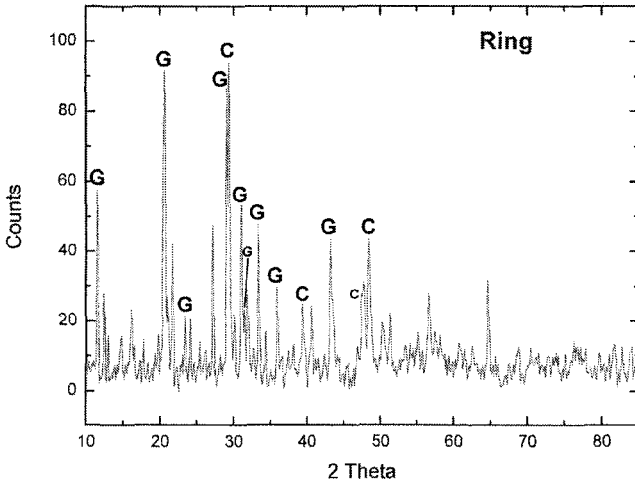
Figure 6. X-ray diffraction spectrum of the white precipitate in the ring at the rock periphery (see Fig. 2). The presence of gypsum ($CaSO_4 \cdot 2H_2O$) in the belt after a brown depleted zone (Figs. 2 and 3) suggests the birth of a first Liesegang ring.

3. Theoretical Modeling

We introduce a simplified reaction scheme adopted from equations 1 and 2, incorporating both dissoltution and precipitation:

$$A + 2X \xrightarrow{k_1} Y \tag{3}$$

$$Y + Z \xrightarrow{k_2} B \tag{4}$$

where A represents a solid mineral (here calcite, $CaCO_3$), X the infiltrating solution (acidic water, H^+), Y the dissolved salt material (Ca^{2+}) from the reaction with X, Z the counter-ion from the infiltrating solution (here SO_4^{2-}) which reacts with Y to form precipitate B ($CaSO_4$).

The coupling of transport (diffusion and percolation flows) to chemical reactions (here dissolution and precipitation) could cause the deposition of the B mineral to take place in the form of bands behind the A dissolution front, just resembling those of a Liesegang pattern. A preliminary illustration of the rock alteration caused by the various reaction-transport processes (adopted from Refs. 3 and 8) is shown in Fig. 7. In 2D, concentric rings of the precipitate are expected to form, originating from a central diffusion source, in a way similar to the behavior observed in Fig. 2 (the marked deviation from circular symmetry is due to the irregular percolation inside the porous medium). In the following treatment, we consider only diffusion (for simplicity) and ignore percolation, convection and water pressure effects. We adopt the model of Müller and Polezhaev[18], modified by Al-Ghoul and Sultan[19,20].

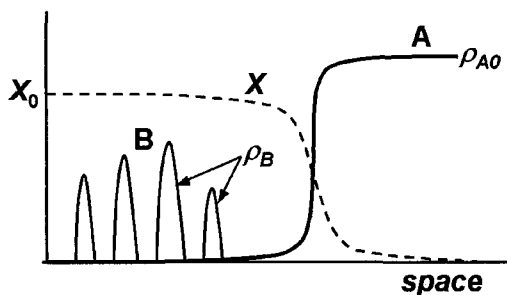The reaction-diffusion processes for the aqueous species X, Y, and Z are de-

Figure 7. Acid infiltration front (X) causing the dissolution of the main rock mineral (A). Another rmineral (B), undergoes a banded precipitation behind the dissolution front.

scribed by equations 5 to 7. D_X, D_Y,and D_Z are the diffusion coefficients, k_1 and k_2 are the rate constants of reactions 3 and 4 respectively, and ∇^2 is the Laplacian operator, which will be reduced to 1D. The $\tilde{\rho}$'s denote the average densities of solid salts in the form of small particles (called nuclei), and the ρ's represent the average density of solid salt in the form of big particles. Let c_A and c_B be the concentrations of the dissolved salts $A\,(aq)$ and $B\,(aq)$ respectively. If c_B exceeds a critical value c_3, small particles (nuclei) nucleate in the salt solution at a rate v_1. If c_B exceeds some other parameter c_2, B nuclei are transformed into large particles at a rate v_3.; otherwise, they dissolve at a rate v_2. When c_B exceeds the value c_1, large particles of the B precipitate grow at a rate v_4.

The evolution equations of ρ_A, ρ_B, $\tilde{\rho}_A$,$\tilde{\rho}_B$, c_A, and c_B are given by (8-13).

The D_C's are the diffusion coefficients of the dissolved salts.

$$\frac{\partial X}{\partial t} = D_X \nabla^2 X - 2k_1 X^2 \rho_A \tag{5}$$

$$\frac{\partial Y}{\partial t} = D_Y \nabla^2 Y + 2k_1 X^2 \rho_A - k_2 Y Z \tag{6}$$

$$\frac{\partial Z}{\partial t} = D_Z \nabla^2 Z - k_2 Y Z \tag{7}$$

$$\frac{\partial c_A}{\partial t} = D_{c_A} \nabla^2 c_A + k_1 X^2 \rho_A - v_1(c_A) + v_2(c_A)\tilde{\rho}_A - v_4(c_A)\rho_A \tag{8}$$

$$\frac{\partial c_B}{\partial t} = D_{c_B} \nabla^2 c_B + k_2 Y Z - v_1(c_B) + v_2(c_B)\tilde{\rho}_B - v_4(c_B)\rho_B \tag{9}$$

$$\frac{\partial \tilde{\rho}_A}{\partial t} = v_1(c_A) - [v_2(c_A) + v_3(c_A)]\tilde{\rho}_A \tag{10}$$

$$\frac{\partial \tilde{\rho}_B}{\partial t} = v_1(c_B) - [v_2(c_B) + v_3(c_B)]\tilde{\rho}_B \tag{11}$$

$$\frac{\partial \rho_A}{\partial t} = v_4(c_A)\rho_A + v_3(c_A)\tilde{\rho}_A - k_1 X^2 \rho_A \tag{12}$$

$$\frac{\partial \rho_B}{\partial t} = v_4(c_B)\rho_B + v_3(c_B)\tilde{\rho}_B + k_2 Y Z \tag{13}$$

The explicit forms of the rate functions are as follows:

$$v_1(c) = \alpha(c - c_3)\Theta[c - c_3] \tag{14}$$

$$v_2(c) = \beta(c_2 - c)\Theta[c_2 - c] \tag{15}$$

$$v_3(c) = \gamma(c - c_2)\Theta[c - c_2] \tag{16}$$

$$v_4(c) = \delta(c - c_1)\Theta[c - c_1] \tag{17}$$

where α, β, γ, and δ are the rate constants and $\Theta(x)$ is the heaviside function. The following initial conditions apply (front located at $x = 0$):

$$\text{At } t = 0, x < 0 \; : \; X = X_0, Y = 0, Z = Z_0 \tag{18}$$
$$\text{At } t = 0, x > 0 \; : \; X = 0, Y = 0, Z = 0$$
$$\text{And } \rho_A(t = 0, x) = \rho_{A_0}; \; \rho_B(t = 0, x) = 0$$

We assume here that once A dissolves, it does not precipitate anymore. Therefore, $\tilde{\rho}_A$ is zero, and ρ_A is expected to decrease. Furthermore, the rate of formation of A nuclei $v_1(c_A)$, and that of growth of particles $v_4(c_A)$ are both zero and hence disappear from the evolution equation for the concentration c_A (equation 8). The latter is thus reduced to:

$$\frac{\partial c_A}{\partial t} = D_{c_A}\nabla^2 c_A + k_1 X^2 \rho_A \tag{19}$$

Similarly equation 12 simplifies to:

$$\frac{\partial \rho_A}{\partial t} = -k_1 X^2 \rho_A \tag{20}$$

Equations 5-13 were solved numerically; the results are plotted in Fig. 8. As expected, the B-mineral exhibits a banded deposition behind the A-dissolution front, in a way similar to the formation of Liesegang bands in the Chemistry laboratory. Thus, the Liesegang banding mechanism could constitute one of many possible routes for the rise of geochemical self-organization. A theoretical study to translate a similar scenario was undertaken by Sultan et. al.[8], but using a different model.

The above study reveals an interesting correlation between the Liesegang phenomenon and the pattern formation in rock systems. The continuation and development of the theoretical work is under close consideration. The experiments require further refinements, but the preliminary results seem to be promising. The time scales of such experiments are very large, and the reactions need to be run for very long times. Also larger rock samples should be used to span larger length scales.
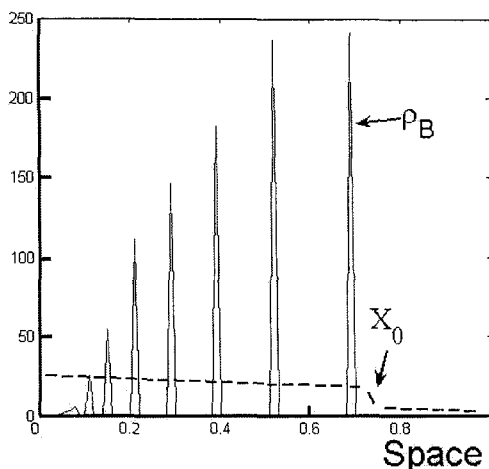
Figure 8. Computer solution of the reaction-diffusion equations 5-13, using the following model parameters: $X_0 = 25; Y_0 = 0; Z_0 = 200; k_1 = 1 \times 10^{-4}; k_2 = 1 \times 10^{-4}; D_X = D_Y = D_Z = 1 \times 10^{-5}; Dc_A = D_{c_B} \ 1 \times 10^{-5}; \rho_{A0} = 20$. The plot (at a fixed late time) shows bands of the B-mineral deposit behind the dissolution front of the embedded A mineral. The dashed line shows the profile of the infiltrating species (acid) X.

Acknowledgments

References

1. R. E. Liesegang, *Chemische Fernwirkung. Lieseg. Photograph. Arch.*, **37**, 305; contin-ued in **37**, 331 (1896).
2. H. K. Henisch, *Crystals in Gels and Liesegang Rings*, Cambridge University Press, Cambridge, 1988.
3. P. Ortoleva, *Geochemical Self-Organization*, Oxford University Press, New York, 1994.
4. H-J. Krug, K-H. Jacob and S. Dietrich, in Ref. [12], pp. 269-282.
5. T. Moxon, *Agate: Microstructure and Possible Origin*, Terra Publications, Doncaster, 1996.
6. T. Moxon, *Education in Chemistry*, **July**, 105 (2001).
7. M. Ohkawa, Y. Yamashita, M. Inoue, R. Itagawa, S. Takeno, *Mineralogy and Petrol-ogy*, **70**, 15 (2000).
8. R. Sultan, P. Ortoleva, F. De Pasquale, and P. Tartaglia, *Earth-Science Reviews*, **29**, 163 (1990).
9. Wi. Ostwald, *Lehrbuch der Allgemeinen Chemie*, **2**. Aufl., Band 2, 2. Teil: Ver-wandtschaftslehre, Engelmann: Leipzig, 1896-1902, p. 778.

236

10. C. Klein, Jr. Hurlbut, *Manual of Mineralogy*, John Wiley and Sons, New York,1993 (21st Ed.).
11. http://onezine.s-one.net.sg/@School/RI/garden.htm.
12. J. H. Kruhl, Ed., *Fractals and Dynamic Systems in Geoscience*, Springer-Verlag, Berlin, 1994.
13. T. Dewers and P. Ortoleva, *J. Phys. Chem.* **93**, 2842 (1989); ibid, *Earth Science Reviews* **29**, 283 (1990).
14. Q. Peng, J.-G. Wu,1 R. D.Soloway, T.-D. Hu, W.-D. Huang,1 Y.-Z. Xu,1 L.-B. Wang,1 X.-F. Li, W.-H. Li, D.-F. Xu and G.-X. Xu, *Biospectroscopy* , **3**(3), 195-205 (1997).
15. D. Xie, J. Wu , G. Xu, Q. Ouyang, R. D. Soloway and T. Hu, *J. Phys. Chem. B* **103**, 8602-8605 (1999).
16. K. Lund, H. S. Fogler and C. C. McCune, *Chem. Eng. Sci.* **28**, 691-700 (1973).
17. K. Lund and H. S. Fogler, *Chem. Eng. Sci.* **31**, 373-380 (1976).
18. S. C. Müller and A. A. Polezhaev, *Chaos*, **4**, 631 (1994).
19. M. Al-Ghoul and R. Sultan, *J. Phys. Chem. A* **105** (34), 8053 (2001).
20. M. Al-Ghoul and R. Sultan, *J. Phys. Chem. A* **107** (8), 1095 (2003).
21. Z. Shrief, M. Al-Ghoul and R. Sultan, *ChemPhysChem* **3**, 592 (2002).

BULK MEDIATED SURFACE DIFFUSION: NON MARKOVIAN AND BIASED BEHAVIOR

JORGE A. REVELLI, CARLOS. E. BUDDE

Facultad de Matemáticas, Astronomía y Física, Universidad Nacional de Córdoba
5000 Córdoba, Argentina

HORACIO S. WIO

Instituto de Física de Cantabria, Universidad de Cantabria and CSIC
E-39005 Santander, Spain
and
Centro Atómico Bariloche and Instituto Balseiro, 8400 Bariloche, Argentina

We analyze the dynamics of adsorbed molecules within the *bulk-mediated surface diffusion* framework. We consider that the particle's desorption mechanism is characterized by a non Markovian process, while the particle's adsorption as well as its motion in the bulk are governed by Markovian dynamics, and include the effect of an external field in the form of a bias normal to the surface. We study this system for the diffusion of particles in a semi-infinite lattice, analyzing the conditional probability to find the system on the reference absorptive plane as well as the surface dispersion as functions of time. The agreement between numerical and analytical asymptotic results is discussed.

During the last few decades, the literature on diffusion processes has covered from diffusion-limited reactions [1] and anomalous diffusion or diffusion in disordered systems [2], through heat conductivity in solids [3] to other diffusion problems in material science [4]. Among the many problems studied in material science, the dynamics of adsorbed molecules at an adsorbing surface is a fundamental issue in interface science and is crucial to a number of emerging technologies [5] (for instance, see Refs. [6,7,8,9] and references therein).

Recently, the mechanism of *bulk-mediated surface diffusion* has been identified and explored. The importance of bulk-surface exchange in relaxing homogeneous surface density perturbations is experimentally well established [10,11,12,13,14,15]. This mechanism typically arises at interfaces separating a liquid bulk phase and a second phase which may be either solid, liquid, or gaseous. Whenever the adsorbed species are soluble in the liquid bulk, adsorption-desorption processes occur continuously. These processes generate surface displacement because desorbed molecules undergo Fickian diffusion in the liquid's bulk, and are latter re-adsorbed elsewhere. When this process is repeated many times, an effective diffusion results for the molecules on the surface.

Usually the studies performed in this type of systems are done within the frame-

work of a Master Equation scheme [16,17,19], where the particle's motion through the bulk and the adsorption-desorption processes are Markovian. In a series of recent papers we have shown some of the most important features of this phenomenon [6,7,8,9]. In particular; we have studied the variance $\langle r^2(t)\rangle_{plane}$ of the position $\vec{r} \equiv (x, y)$ on the interface, and the conditional probability $P(z = 1, t) \equiv \sum_{x,y} P(x, y, z = 1; t|0, 0, 1; t = 0)$ to find the particle on the interface at time t, if it was initially at (0,0,1), by means of analytical and numerical methods. In those papers we have studied the cases of infinite and bounded bulk, where the particle dynamics is governed by Markovian and non Markovian processes.

In this work we address the dynamics of adsorbed molecules when, in addition to the particle's desorption mechanism being characterized by a non Markovian process –while the particle's adsorption and its motion in the bulk are governed by a Markovian dynamics–, there is a biased behavior for the motion along the vertical axis (that is, we consider the effect of some external field normal to the interface). Non Markovian desorption process can occur when the surfaces contains "deep traps", capture and re-emission from surfaces that contain sites with several internal states such as the "ladder trapping model", proteins with active sites deep inside its matrix, etc. [20,21]. When non Markovian process are present, it is necessary to resort to *Generalized Master Equations*, that are characterized by a "memory kernel" and may be related univocally with a CTRW scheme [22].

Our aim in this work is to study the influence of both, the non Markovian desorption together with the indicated biased dynamics, on the effective diffusion process at the interface $z = 1$. For that purpose we calculate the temporal evolutions of the $\langle r^2(t)\rangle_{plane}$ and of $P(z = 1, t)$.

Here we exploit the same framework as in [8,9]. $P(n, m, l; t)$ satisfies

$$
\dot{P}(n, m, 1; t) = \gamma_1 P(n, m, 2; t) - \int_0^t dt' \, K(t') P(n, m, 1; t - t'),
$$
$$
+\alpha_1 [P(n - 1, m, 1; t) + P(n + 1, m, 1; t) - 2P(n, m, 1; t)]
$$
$$
+\beta_1 [P(n, m - 1, 1; t) + P(n, m + 1, 1; t) - 2P(n, m, 1; t)], \quad \text{for } l = 1
$$
$$
\dot{P}(n, m, 2; t) = \int_0^t dt' \, K(t') P(n, m, 1; t - t') + \gamma_1 P(n, m, 3; t)
$$
$$
-\gamma_2 P(n, m, 2; t) - \gamma_1 P(n, m, 2; t)
$$
$$
+\alpha [P(n - 1, m, 2; t) + P(n + 1, m, 2; t) - 2P(n, m, 2; t)]
$$
$$
+\beta [P(n, m - 1, 2; t) + P(n, m + 1, 2; t) - 2P(n, m, 2; t)], \quad \text{for } l = 2
$$
$$
\dot{P}(n, m, l; t) = \alpha [P(n - 1, m, l; t) + P(n + 1, m, l; t) - 2P(n, m, l; t)]
$$
$$
+\beta [P(n, m - 1, l; t) + P(n, m + 1, l; t) - 2P(n, m, l; t)]
$$
$$
+\gamma_1 [P(n, m, l + 1; t) - P(n, m, l; t)] + \gamma_2 [P(n, m, l - 1; t)
$$
$$
-P(n, m, l; t)], \qquad\qquad \text{for } l \geq 3 \qquad (1)
$$

where α and β are the transition probabilities per unit time through the bulk in the x, y directions respectively; while γ_1 and γ_2 are the transition probabilities per

unit time through the bulk in the z direction that, when $\gamma_1 \neq \gamma_2$ indicates a biased behavior (or the presence of an external field). It is important to note that the model presented in Eq. (1), allows for the possibility that the particles can move in the plane $z = 1$ with temporal frequencies α_1 in the x direction and β_1 in the y direction. $K(t)$ represents the memory kernel for desorption at all sites over the $z = 1$ surface (that is for $(n, m, l = 1)$). It is clear that, when $\gamma_1 = \gamma_2 = \gamma$ and $\alpha_1 = \beta_1 = 0$, we recover the same set of equations used in [8,9].

Taking the Fourier transform with respect to the x and y variables and the Laplace transform with respect to the time t in the above equations, we obtain

$$
\begin{aligned}
sG(k_x, k_y, 1; s) - P(k_x, k_y, 1, t = 0) &= \gamma_1 G(k_x, k_y, 2; s) + A_1(k_x, k_y)G(k_x, k_y, 1; s) \\
&\quad -K(s)G(k_x, k_y, 1; s), \qquad\qquad \text{for } l = 1 \\
sG(k_x, k_y, 2; s) - P(k_x, k_y, 2, t = 0) &= A(k_x, k_y)G(k_x, k_y, 2; s) + K(s)G(k_x, k_y, 1; s) \\
&\quad -\gamma_1 G(k_x, k_y, 2; s) + \gamma_1 G(k_x, k_y, 3; s) \\
&\quad -\gamma_2 G(k_x, k_y, 2; s), \qquad\qquad \text{for } l = 2 \\
sG(k_x, k_y, l; s) - P(k_x, k_y, l, t = 0) &= A(k_x, k_y)G(k_x, k_y, l; s) + \gamma_1[G(k_x, k_y, l - 1; s) \\
&\quad -G(k_x, k_y, l; s)] + \gamma_2[G(k_x, k_y, l + 1; s) - G(k_x, k_y, l; s)], \\
&\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{for } l \geq 3 \qquad (2)
\end{aligned}
$$

Here we have used the same definitions as in [6,7,8,9]. In particular we have $A(k_x, k_y) = 2\alpha[\cos(k_x) - 1] + 2\beta[\cos(k_y) - 1]$, and $A_1(k_x, k_y) = 2\alpha_1[\cos(k_x) - 1] + 2\beta_1[\cos(k_y) - 1]$.

Clearly, the above equations for $G(k_x, k_y, l; s)]$ are similar to Eq. (2) in [7], with $K(s)$ instead of δ, and the unbiased γ replaced by γ_1 and γ_2. Therefore all results obtained in [6,7] remains valid for a non Markovian dynamics when δ is replaced by $K(s)$, and the bias is adequately included.

In what follows, and in order to simplify and to focus on the joint effect of both, the non Markovian desorption and the bias, we adopt $\alpha = \beta$ and $\alpha_1 = \beta_1$. By using Eq. (12) of Ref. [7], $\langle r^2(s) \rangle_{plane}$ turns out to be the ratio of two complicated functions of s

$$
\langle r^2(s) \rangle_{plane} = \frac{N(s)}{D(s)}, \qquad (3)
$$

where

$$N(s) = (8\alpha K(s)\gamma_1)\left(\gamma_1^3 + \gamma_2^3 + s^3 + 3\gamma_1^2 s + 3\gamma_2^2 s + 3\gamma_1\gamma_2 s\right.$$

$$\left. + 3\gamma_1 s^2 + 3\gamma_2 s^2 - \sqrt{(\gamma_1 + \gamma_2 + s)^2 - 4\gamma_1\gamma_2}(\gamma_1^2 + (\gamma_2 + s)^2 + \gamma_1(\gamma_2 + 2s))\right)$$

$$+ (8\alpha_1\gamma_1\gamma_2)\left(\sqrt{(\gamma_1 + \gamma_2 + s)^2 - 4\gamma_1\gamma_2}(\gamma_1^2 + 2\gamma_1 s + (\gamma_2 + s)^2)\right.$$

$$\left. - (\gamma_1 + \gamma_2 + s)((\gamma_1 - \gamma_2)^2 + 2(\gamma_1 + \gamma_2)s + s^2)\right), \qquad (4)$$

and

$$D(s) = \sqrt{(\gamma_1 + \gamma_2 + s)^2 - 4\gamma_1\gamma_2}\left\{\sqrt{\frac{\gamma_1}{\gamma_2}}((\gamma_2 - K(s))\right.$$

$$\left((\gamma_2 - K(s))(\gamma_1^2 + 2\gamma_1 s + (\gamma_2 + s)^2)\right.$$

$$\left. - \sqrt{(\gamma_1 + \gamma_2 + s)^2 - 4\gamma_1\gamma_2}(\gamma_2 - K(s))(\gamma_1 + \gamma_2 + s)\right)$$

$$+ \sqrt{\gamma_1\gamma_2}\left(\sqrt{(\gamma_1 + \gamma_2 + s)^2 - 4\gamma_1\gamma_2}(\gamma_1 + \gamma_2 - K(s))\right.$$

$$\left.\left. - (\gamma_1^2 + (\gamma_2 - K(s))(\gamma_2 + s)) + \gamma_1(s - K(s))\right)\right\}^2 \qquad (5)$$

For $P(z = 1, t)$, according to the result shown in [6], we have

$$\tilde{P}(z = 1, s) = 2\gamma_1\gamma_2$$

$$\left(2\gamma_1\gamma_2 s + K(s)[\gamma_1\gamma_2 + \gamma_1(\sqrt{(\gamma_1 + \gamma_2 + s)^2 - 4\gamma_1\gamma_2} - s - \gamma_1)]\right)^{-1} \quad (6)$$

Finally, the relation between $\psi(t)$, the waiting time density, as defined in a CTRW scheme, and $K(t)$, the memory kernel in Eq (1) [22], in the Laplace domain reads $K(s) = s\,\psi(s)\,[1 - \psi(s)]^{-1}$.

We show the results, obtained via Tauberian theorems [18], for the asymptotic long time system's behavior in two cases: when the waiting time function has a short or a long time tail. We assume that, when $s \ll 1$, $\psi(s) \sim 1 - Bs^\nu$ with $0 < \nu < 1$ for the case of a *long tail*, and $\nu = 1$ for the *short tail case*. In the case of short tail waiting time densities we have $B = < t >_{ads}$. Consequently, in this limit $K(s) \sim \frac{1}{B}s^{1-\nu}$.

Considering the different possibilities we have three cases: (i) $\gamma_1 > \gamma_2$; (ii) $\gamma_1 < \gamma_2$; (iii) $\gamma_1 = \gamma_2 = \gamma$.

Short Tail

- $\gamma_1 > \gamma_2$

$$< r^2(t) >_{plane} \sim \left[\frac{4(\gamma_1 - \gamma_2)}{(\gamma_1 - \gamma_2 + \frac{1}{B})^2} \left(\frac{\alpha}{B} + \alpha_1(\gamma_1 - \gamma_2) \right) \right] t, \qquad (7)$$

and

$$P(z = 1, t) \sim \left\{ 1 + \frac{1}{B(\gamma_1 - \gamma_2)} \right\}^{-1} \qquad (8)$$

- $\gamma_1 < \gamma_2$: In this case, both $< r^2(t) >_{plane}$ and $P(z = 1, t)$, show a time exponential decay.
- $\gamma_1 = \gamma_2 = \gamma$

$$< r^2(t) >_{plane} \sim 4 B \frac{\alpha \gamma^{\frac{1}{2}}}{\pi^{\frac{1}{2}}} t^{\frac{1}{2}}, \qquad (9)$$

and

$$P(z = 1, t) \sim \frac{B \gamma^{\frac{1}{2}}}{\pi^{\frac{1}{2}}} t^{-\frac{1}{2}}, \qquad (10)$$

Long Tail

- $\gamma_1 > \gamma_2$

$$< r^2(t) >_{plane} \sim \frac{4\alpha_1(\gamma_1 - \gamma_2)^2}{\gamma_2^2} t + 4\alpha_1 \frac{(\gamma_1 - \gamma_2)}{B\gamma_2\Gamma(1 + \nu)} t^\nu, \qquad (11)$$

and

$$P(z = 1, t) \sim 1. \qquad (12)$$

- $\gamma_1 < \gamma_2$

$$< r^2(t) >_{plane} \sim \frac{4(B\gamma_1\gamma_2)^2 \alpha_1}{(\gamma_1 - \gamma_2)^2 \Gamma(2(1 - \nu))} t^{1-2\nu} + 4B\gamma_2\gamma_1^2\alpha_1 \frac{(\gamma_1 - \gamma_2)^3}{\Gamma(\frac{5}{2} - \nu)} t^{-\nu}, \qquad (13)$$

and

$$P(z = 1, t) \sim \frac{B\gamma_2}{\Gamma(1 - \nu)(\gamma_2 - \gamma_1)} t^{-\nu}, \qquad (14)$$

- $\gamma_1 = \gamma_2 = \gamma$ Here we have different asymptotic behaviors depending on the range of values of ν.

$$< r^2(t) >_{plane} \sim \frac{(2\alpha)}{B\sqrt{\gamma}\Gamma(\frac{3}{2} + \nu)} t^{\frac{1}{2}+\nu} + 4\alpha_1 t, \qquad 0 < \nu < \tfrac{1}{2}$$

$$< r^2(t) >_{plane} \sim \frac{(2\alpha)B\sqrt{\gamma} + 4\gamma B^2\alpha_1}{(B\sqrt{\gamma} + 1)^2} t + 2 B \frac{\sqrt{\gamma}}{\Gamma(\frac{5}{2} - \nu)} t^{\frac{3}{2}-\nu}, \qquad \nu = \tfrac{1}{2}$$

$$< r^2(t) >_{plane} \sim \frac{4\gamma B^2\alpha_1}{\Gamma(3 - 2\nu)} t^{2(1-\nu)}, \qquad \tfrac{1}{2} < \nu < 1, \qquad (15)$$

242

and

$$P(z = 1, t) \sim 1, \qquad\qquad\qquad 0 < \nu < \tfrac{1}{2}$$

$$P(z = 1, t) \sim \frac{B\sqrt{\gamma}}{(B\sqrt{\gamma} + 1)}, \qquad\qquad \nu = \tfrac{1}{2}$$

$$P(z = 1, t) \sim \frac{B\sqrt{\gamma}}{\Gamma(\tfrac{1}{2} - \nu)} t^{\frac{1}{2} - \nu}, \qquad\qquad \tfrac{1}{2} < \nu < 1, \qquad (16)$$
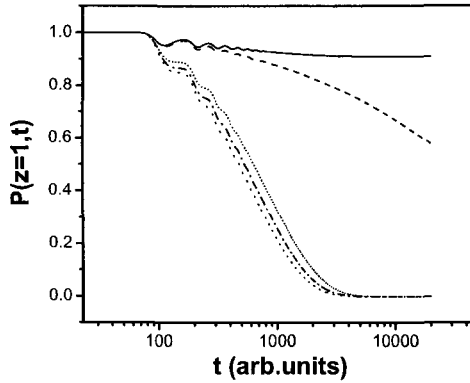

Figure 1. Temporal evolution of $P(z = 1, t)$, the probability distribution on the plane $z = 1$. Here we have adopted $\alpha = \beta = \alpha_1 = \beta_1 = 1$, $\theta = 0.01$, $a = 50$ and $\gamma_1 = 1$. The different curves correspond to $\gamma_2 = 0.9$ (continuous line), $\gamma_2 = 1.0$ (dashed line), $\gamma_2 = 3.0$ (dotted line), $\gamma_2 = 5.0$ (dashed-dotted line).

Now we show some of the results obtained numerically. As was discussed in previous works [6,7], for general cases, the Laplace transform of $\langle r^2(s) \rangle_{plane}$ and $P(z = 1, s)$ usually cannot be analytically inverted. As this occurs in the present case, we have been forced to employ a numerical inversion method. The efficacy of such a method was established in [6,7], where it was shown that it is a reliable tool and that we can trust their results in those cases where analytical results are not accessible [23].

To describe the desorption dynamics from the surface we have used the following family of waiting time densities $(\psi(t))$ that has been extensively exploited in modeling non Markovian situations [28,18]. The form is

$$\psi(t) = \theta a \frac{(\theta a t)^{(a-1)}}{\Gamma(a)} e^{-\theta a t}, \qquad (17)$$

where a is a positive integer ($a = 1, 2, 3, ..$) and $\Gamma(a)$ is the Gamma function. There are two parameters which characterized the function: a, called *markovianicity parameter*, defines the degree of non markovianicity of the function ($a = 1$ corresponds

to the Markovian case, $a \neq 1$ to the non-Markovian case); while the parameter θ is the "average desorption's rate". Also, as shown in [24], the mean value of these waiting time densities is $< t > = \int_0^\infty t \; \psi(t) \; dt = \theta^{-1}$. Hence, the "average desorption's time" does not depends on the a parameter, but is only function of the desorption's rate [24].
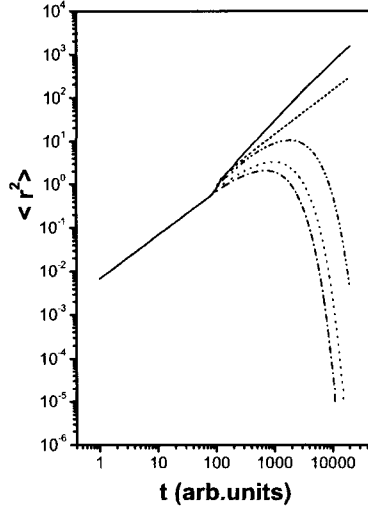

Figure 2. Temporal evolution of $\langle r^2(t) \rangle_{plane}$, the variance on the plane $z = 1$. Here we have adopted $\alpha = \beta = 1$, $\alpha_1 = \beta_1 = 0.01$, $\theta = 0.01$, $a = 50$ and $\gamma_1 = 1$. The different curves correspond, from top to bottom $\gamma_2 = 0.2$, $\gamma_2 = 1.$, $\gamma_2 = 1.2$, $\gamma_2 = 1.5$, $\gamma_2 = 1.7$, $\gamma_2 = 3.$, $\gamma_2 = 5.$.

In Fig. 1 we show the temporal evolution for $P(z = 1, t)$ when $\alpha = \beta = \alpha_1 = \beta_1 = 1$, $\theta = 0.01$, $a = 50$ and $\gamma_1 = 1$, for different values of γ_2. It is clear the agreement between this numerical results and the asymptotic behavior indicated above, as well as in the case of Fig. 2 where we show the temporal evolution for $\langle r^2(t) \rangle_{plane}$.

In order to analyze the long time tail case, we have consider the following desorption waiting time density function (defined in the Laplace domain) [29]

$$\psi(s) = \frac{1}{1 + (\frac{s}{\phi})^\nu}, \qquad 0 < \nu < 1 \tag{18}$$

However, as ϕ amounts to only a change in the time scale, we adopted $\phi = 1$. We have also fixed the parameters α, β, α_1, β_1 and γ_1, all equal to one.

In Fig. 3 we show the temporal evolution for $P(z = 1, t)$, while in Fig. 4 we show $\langle r^2(t) \rangle_{plane}$. Again the agreement with the asymptotic behavior indicated above is
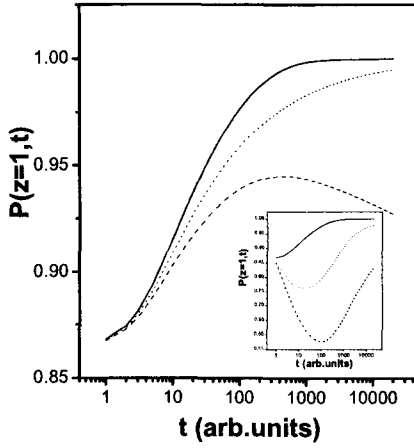
apparent.

Figure 3. Temporal evolution of $P(z = 1, t)$, the probability distribution on the plane $z = 1$. Here we have adopted $\alpha = \beta = \alpha_1 = \beta_1 = 1$, $\phi = 1$. and $\gamma_1 = 1$, while we adopt $\nu = 0.1$. The different curves correspond to $\gamma_2 = 0.8$ (continuous line), $\gamma_2 = 1.0$ (dotted line), $\gamma_2 = 1.1$ (dashed-dot line), $\gamma_2 = 1.2$ (dashed line). The inset shows the temporal evolution of $P(z = 1, t)$, the probability distribution on the plane $z = 1$. Here we have adopted $\alpha = \beta = \alpha_1 = \beta_1 = 1$, $\phi = 1.$, $\gamma_1 = 1$ and $\gamma_2 = 0.8$. The different curves correspond to $\nu = 0.1$ (continuous line), $\nu = 0.5$ (dashed line), $\nu = 0.7$ (dotted line).

Summarizing, we have studied the evolution of particles making an effective diffusion on a surface. The diffusion is actually performed both, on the surface, and across the bulk surrounding the surface, resulting in the so called *bulk mediated surface diffusion* phenomenon. The main feature of this work was to present an analytical model for non Markovian desorption from the surface also including the effect of an external (normal to the surface) field, through a biased behavior in the normal direction. For the bulk that surrounds the surface we have considered that it is semi-infinite, and that the particles undergoes a Markovian motion there.

We observed the influence of the both effects on the effective diffusion process at the interface $z = 1$ by calculating analytically, in the Laplace domain, the temporal evolutions of $\langle r^2(t) \rangle_{plane}$, the variance, and $P(z = 1, t)$, the conditional probability of being on the surface at time t since the particle arrived there at $t = 0$. We have chosen two families of non Markovian desorption waiting time densities and analyzed the numerical evaluation for $\langle r^2(t) \rangle_{plane}$ and $P(z = 1, t)$. One desorption waiting time densities with a **finite first moment**, while the other corresponds to a waiting time density for desorption has an *infinite* **first moment**. We have been able to deduce the asymptotic behavior in both cases, determining regions of values of the parameter ν, obtaining asymptotic **sub-diffusive** regimes, or cases where **normal diffusion** takes place. The numerical results are in agreement with
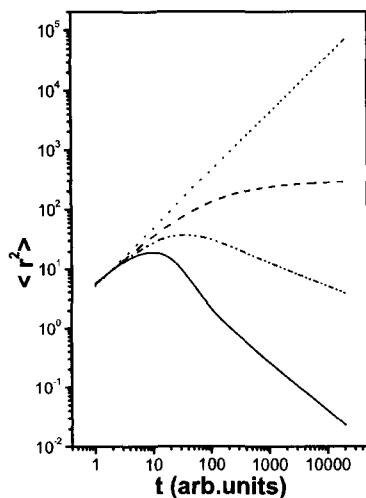
the indicated asymptotic ones.

Figure 4. Temporal evolution of $\langle r^2(t)\rangle_{plane}$, the variance of the position $\vec{r} \equiv (x, y)$ on the plane $z = 1$. Here we have adopted $\alpha = \beta = \alpha_1 = \beta_1 = 1$, $\gamma_1 = 1$ and $\gamma_2 = 10$. The different curves correspond to $\nu = 0.1$ (continuous line), $\nu = 0.5$ (dashed line), $\nu = 0.7$ (dashed double dot line), $\nu = 0.9$ (dotted line).

Finally, it is worth remarking here an important aspect of the present approach. Through the above results we have shown that the behavior of $\langle r^2(t)\rangle_{plane}$ and of $P(z = 1, t)$ are strongly dependent on both, the desorption mechanism and the effect of bias. As the effective dispersion and the percentage of particles that remain in $z = 1$ are measurable magnitudes [17] they may be used to investigate the characteristic and fundamental parameters of the desorption processes.

Acknowledgments

References

1. S.A. Rice, *Diffusion-Limited Reactions*, (Elsevier, Amsterdam, 1985);
   A.A. Ovchinnicov, S.F. Timashev and A.A. Belyy, *Kinetics of Diffusion Controlled Chemical Physics*, (Nova Science, Commack, New York, 1989).

2. J.P. Bouchaud and A. Georges, Phys. Rep. **195** 127 (1990).
3. H.S. Carslaw and J.C. Jaeger, *Conduction of Heat in Solids*, (Oxford University Press, London, 1959).
4. M. Doi and S.F. Edwards, *The Theory of Polymer Dynamics*, (Calrendon, Oxford, 1986);
   Y. Fukai, *The Metal-Hydrogen System*, (Springer, Berlin, 1993).
5. J. H. Clint, *Surfactant Aggregation* (Chapman and Hall, New York, 1992).
6. J.A. Revelli, C.E. Budde, D. Prato and H.S. Wio, Europ. Phys. J. B **36**, 245-251 (2003).
7. J.A. Revelli, C.E. Budde, D. Prato and H.S. Wio, Europ. Phys. J. B **37**, 205 (2004).
8. J.A. Revelli, C.E. Budde, D. Prato and H.S. Wio, Europ. Phys. J. B **43**, 65 (2005).
9. J.A. Revelli, C.E. Budde, D. Prato and H.S. Wio, New J. of Phys. **7**, 16 (2005), http://stacks.iop.org/1367-2630/7/16.
10. C.T. Shibata and A. M Lenhof, J. Colloid Interface Sci. **148**, 469 (1992), and **148**, 485 (1992).
11. V. G. Levich, *Physiochemical Hydrodynamics* (Prentice-Hall, Inc. Englewood Clliffs NJ. 1962), 2nd ed.
12. E. H. Lucassen-Reynders and J. Lucassen, Adv. Colloid Interface Sci **2** 347 (1969).
13. E. H. Lucassen-Reynders, J. Lucassen P. R. Garret D. Giles and F. Hollway, Adv. Chem. Scr. **144**, 272 (1975).
14. S-Y. Lin K. McKeigue, and C. Maldarelli, AIChE J. **36** (12), 1785 (1990).
15. M. Coppey, O. Bénichou, J. Klafter, M. Moreau, and G. Oshanin, Phys. Rev. E **69**, 036115 (2004).
16. O. V. Bychuk and B. O' Shaughnessy, Phys. Rev. Lett. **74**, 1795 (1995).
17. O. Bychuk and B. OShaughnessy, J. Chem. Phys., **101**, 772 (1994);
   S. Stapf, R. Kimmich and R. O. Seitter, Phys. Rev. Lett., **75**, 2855 (1995).
18. E.W. Montroll and B.J. West, in *Fluctuation Phenomena*, E.W. Montroll and J.L. Lebowitz, eds. (North Holland, Amsterdam, 1979).
19. N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 2nd ed. (North Holland, Amsterdam, 1993).
20. J. W. Haus and K. W. Kehr . Phys Rep. **150**, 265 (1987).
21. M. A. Re and C. E. Budde. Phys. Rev. E. **61**, 1110 (2000).
22. M.F. Shlesinger and G.H. Weiss, Eds., *The Wonderfoul World of Stochastics*, (North-Holland, Amsterdam, 1994).
23. G. Honig and U. Hirdes, *On the Application of an Efficient Algorithm for the Numerical Laplace Inversion*, Kernforschungsanlage Jülich GmbH, Institut für Festkörperforschung (August 1980);
   J. A. Revelli, Ph.D. Thesis, Instituto Balseiro, Universidad Nacional de Cuyo, Argentina (2003).
24. J.A. Revelli, C.E. Budde and H.S. Wio, Phys. Lett. A **306**, 104 (2002).
25. G. Pfister, H. Scher, Adv. Phys. **27**, 74 (1978).
26. J. Noolandi, Phys. Rev. B **16**, 4466 (1977)
27. D. Richter, T. Springer, Phys. Rev. **B18**, 126 (1978)
28. H. Scher and M. Lax, Phys. Rev. B **7**, 4491 and 4502 (1973).
29. G. H. Weiss, Adv. Chem. Phys **13**, 1 (1966);
   G.H. Weiss, in *Aspects and Applications of the Random Walk*, (North Holland, Amsterdam, 1994).

MULTIFRACTAL FORMALISM FOR REMOTE SENSING: A CONTRIBUTION TO THE DESCRIPTION AND THE UNDERSTANDING OF METEOROLOGICAL PHENOMENA IN SATELLITE IMAGES

J. GRAZZINI [*]

*RAD - IACM (FORTH), Vassilika Vouton, P.O. Box 1527,
71110 Heraklion, Greece. grazzini@iacm.forth.gr*

A. TURIEL [†]

*ICM - CMIMA (CSIC), Passeig Marítim de la Barceloneta 37-49,
08003 Barcelona, Spain. turiel@icm.csic.es*

H. YAHIA

*CLIME - INRIA Rocquencourt, Domaine de Voluceau, B.P. 105,
78153 Le Chesnay cedex, France. hussein.yahia@inria.fr*

The ability of fractals to mimic Nature has led to the widespread acceptance of fractal, and, beyond, multifractal models. Such models are in the roots of new approaches in environmental sciences for processing images displaying turbulent-like systems. This paper addresses the problem of detecting critical areas associated with convection in satellite meteorological images. The technique we propose takes information about the spatial domain and relies on general statistical concepts. Due to the turbulent character of the observed atmospheric systems, the multifractal approach is naturaly adopted herein to describe not only the geometrical properties of images but also the underlying physical phenomena involved. The multifractal formalism leads first to the classification of different chaotic parts of systems according to their dynamical significance. It is further employed to extract the places at which convection takes place in the flow. It is shown that it finally allows the determination of information that would be otherwise hidden. Without any temporal information, this remote sensing technique has potential application to infer the convective-scale processes occurring in individual convective systems. More generally, it leads to new insights into the analysis of natural phenomena from still images.

1. Introduction

It is now recognized that many processes in nature, engineering, science and economics exhibit complex scaling behavior [15]: some obvious examples are given by oceanographic and atmospheric phenomena [6,13], other less evidently related systems

---

are finance markets [16], topographic landscapes [11] or even heartbeat rythms [12]. This complexity in scaling is commonly reflected as a nonlinear irregular variability of patterns fluctuations across a broad range of scales [20,21]. Theoretical arguments have shown that turbulence is representative of such processes [2,7]. This has led to a growing interest in obtaining compact description and accurate prediction tools for turbulent-like systems. However, the analysis of turbulent experimental data is still regarded as strongly challenging: classical descriptions are based on Navier-Stokes's or other chaos-spawning differential equations, for which inverse and direct problems are extremely demanding; they do not supply with a satisfactory quantification of the nonlinear variability in multiple scaling [15]. To overcome these limitations, the concept of *multifractals* has been developed to provide a mathematical model that embraces the irregularities that are, more generally, generic to the evolution of dynamical systems in Nature [15,18]. The multifractal formalism mainly assumes that local singularities exist within the systems and are arranged in a complicated mesh of different scale-invariant (fractal) structures. Detailed discussions on the multifractal nature of fluid turbulence can be found in the literature [6,7,20].

In this context, quantification of atmospheric flow patterns as recorded by meteorological parameters, besides helping prediction of weather and climate, also provides a model for turbulent fluid flows. In particular, the improvement in the latest years of the technology for observing the Earth from the space has given rise to a growing interest in the developpement of new signal processing tools derived from thermodynamical concepts and their application on satellite images [3,5]. Indeed, for most of the large-scale processes, statistics can only be obtained from satellite observation - and especially infrared (IR) images - because only this kind of data can cover the range of time and space scales involved in atmospheric phenomena [1,10]. The objective of the present work is to show how the multifractal scaling formalism can be: *(i)* applied to process satellite images in order to extract meaningful meteorological patterns; *(ii)* further extended in order to give an interpretation to these patterns with regard to the underlying geophysical phenomena. We precisely focus on the characterization of the dynamics of convective systems (CS) in satellite IR images. Such knowledge is of great interest for a number of climate applications - such as general circulation models - as CS are mainly responsible for hard weather situations like rainfalls and thunderstorms [1]. However, the analysis of their life-cycle is particularly challenging due to their multifractal, intermittent structure: evidence of scaling in space and time of clouds has been demonstrated in many previous studies [5,9,13]. Multifractal geometry appears naturally as the suitable approach to perform such analysis, as we expect any quantity defined on the atmospheric flow to behave in an ergodic way, and even more to define a multifractal structure [2,20]. Indeed, it has already provided various tools for the investigation of cloud properties: multifractal models are used [3] to analyze the textural roughness of clouds; the multifractal cascade phenomenology is adopted [5] to identify convective-scale processes in CS; it is also used [13] to classify measured

rain fields with respect to the nature of the observed clouds; multifractal temporal fluctuations associated with selfsimilar multifractal spatial patterns are used [20] for rain areas detection; multifractal analysis of singularities is performed [8] to extract the main thermal fronts of CS. Recently, a new method for the assessment and tracking of pluviometry in CS in still IR images has been proposed [23]. The main underlying idea [8] is that strong transitions in the signal are related with the dynamics of the flow. The streamlines of IR data are first identified by a multifractal singularity analysis; a proxy image that simulates pure horizontal advection is then derived from these streamlines; finally, from the comparison of both original and proxy images, convective areas of the flow are localized, and further identified with purely raining places. It appears that the analysis of the flow can be reduced to the characterization of some singular points derived from the multifractal formalism. In this paper, based on the assertion that forms visualized as still images account for the motion, we develop and evaluate this approach to the problem of motion characterization. The multifractal formalism [23] allows to identify the "hidden sources" for the process, providing critical information about the flow evolution and the nature of the underlying motion: these sources correspond to areas where the (2D) image motion field is divergent, or, equivalently, where the (3D) flow is convective.

The paper is organized as follows. In section 2, we review the multifractal model for extracting meaningful patterns in IR images and for reconstructing a proxy image. In section 3, we relate this model with the hypothesis describing advective motion and we develop the formalism of the sources. In section 4, we compare the results with a physically-based optical-flow model and we provide an interpretation to the sources regarding the underlying phenomena. As a conclusion, we give the natural continuation to this work for further validation/use of the model.

2. Multifractality help to decipher complex image patterns

Multifractality is a property of turbulent-like signals which is present in very different physical systems [7]. It is generally reported on intensive, scalar variables of chaotic structures. In the following, we present one of the possible formalisms used to characterize multifractality, that was introduced in [24], and apply it on IR images.

The multifractal structure of a signal I can be assessed by defining at any particular point $\vec{x}$ a measure μ in the way [24,25]:

$$\mu(B_r(\vec{x})) \equiv \int_{B_r(\vec{x})} d\vec{y} \, |\nabla I|(\vec{y}) \, , \tag{1}$$

where the norm of the gradient $|\nabla I|$ is weighted over all the points in the ball $B_r(\vec{x})$ of radius r around $\vec{x}$, and observing local power-law scalings [a]:

$$\mu(B_r(\vec{x})) \approx \alpha(\vec{x}) \, r^{h(\vec{x})} \, . \tag{2}$$

---

[a]Note that the exponents $h(\vec{x})$ in eq. (2) could be rewritten $h(\vec{x}) + d$ where d is the dimension of the space so that they will be the same for measures of the same degree of regularity disregarding the dimensionality [25].
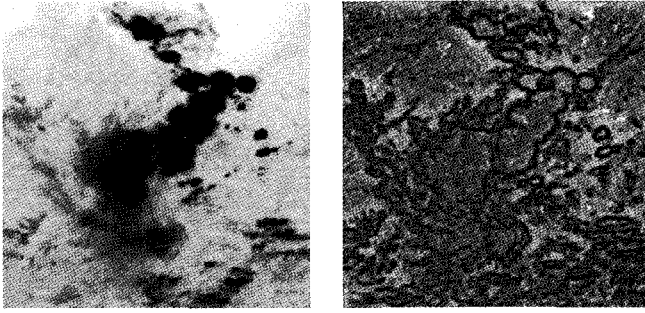
Figure 1.: Left: IR image acquired by the geostationnary satellite Meteosat displaying a typical CS. Right: representation of singularity exponents; the darker the pixel, the lower the value of this exponent at that point. Lower values (*i.e.* most singular points) are found near the boundaries.

That power-law behaviour implies that there is no privileged scale of observation: the signal is self-similar as all the dependence in the scale r is contained in the factor $r^{h(\vec{x})}$, not $\alpha(\vec{x})$. Let then F_{h_0} be the level-set associated to a given value h_0:

$$F_{h_0} \equiv \{\vec{x} \ : \ h(\vec{x}) = h_0\}, \tag{3}$$

hence, all points $\vec{x}$ in the space are arranged according to the local *singularity exponent* $h(\vec{x})$ they are assigned. The family of subsets F_h are of fractal character (*i.e.* they exhibit the same geometrical structure at different scales) due to the scale invariance of the function h and provide a multifractal splitting of the image. Theoretical foundations are fully discussed elsewhere [25]. Besides, this decomposition has a strong dynamical meaning: one can deduce the statistics of changes in scale just knowing the dimensions of the fractal components [18]. Let notice that, in turbulence theory, an expression similar to eq. (1) defines the local dissipation of energy in the flow; thus, the decomposition F_h can be regarded as the geometrical representation of the multifractal cascade of energy between the different scales [2,9]. However, the current approach does not depend on the specifics of the cascade phenomenology, the geometrical attributes of the multifractal hierarchy are analyzed directly.

In the context of image processing, the point $\vec{x}$ refers to the pixel, the variable $I(\vec{x})$ is the graylevel value and the measure μ expresses the local variability of graylevels around $\vec{x}$. Performing a multifractal decomposition with eq. (2) finally amounts to locally quantify the strenght of the transition the signal undergoes around each pixel. In practice [25], to overcome the difficulties due to the image discretization, the gradient $I(\vec{x})$ is simply estimated by finite difference at the distance of one resolution box (one pixel)[b]. The singularity exponents are computed through a log-log linear regression on the wavelet transform [14] of the measure, which has been proven to also exhibit multifractal behaviour [25], and not of the measure itself.

---

[b]This definition is derived from the linear increment formalism originally employed for the analysis of multifractal behavior in turbulent flows [7].
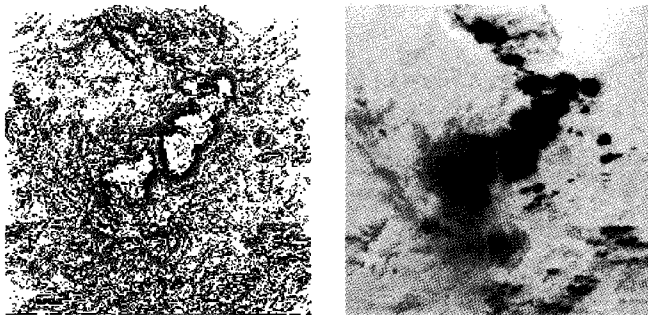
Figure 2.: Left: MSM estimated on the sample IR image of Fig. 1; pixels of the MSM are assigned a non-white graylevel according to the orientation of the gradient over this set (grey: negative; black: positive). Right: reconstruction out of the data defined on the orientated MSM.

When considering meteorological IR images (Fig. 1, left), the graylevel intensity is identified with thermal IR radiance, measured on higher layers of the atmosphere, which is a typical turbulent variable following a complicated pattern. Namely, it was shown [8] that IR images are of multifractal nature. In Fig. 1, right, we can see a typical segmentation induced by the spatial distribution of the singularity exponents. The F_h splitting consists of an edge-like most singular manifold surrounded by manifolds of decreasing singularity as we move away from it [25].

In the multifractal decomposition, one set, called the *Most Singular Manifold* (MSM), is of particular importance: the manifold F_∞ associated to the minimum value for the singularities $h_\infty \leq h(\vec{x}) \, \forall \, \vec{x}$, whose existence is ensured by the multifractal theory [24]. The MSM has usually a fractal co-dimension of 1 and resembles the edges or contours of the objects in the case of natural images [25] or relevant dynamical points in other systems [8]. In Fig. 2, left, the MSM gathers the pixels where the discontinuities in graylevel are the more appreciable, thermal fronts indeed. The MSM is known as the most informative set in the image [8], but its most relevant property is found in its interpretation as the origin of the multifractal hierarchy [22]. It is possible (under certain statistical hypothesis) to reconstruct the multifractal completely from the values of the gradient restricted to the MSM, ∇I_{h_∞}:

$$I(\vec{x}) \; = \; \vec{g} \otimes \nabla I_{h_\infty}(\vec{x}) \tag{4}$$

with a particular reconstructing kernel $\vec{g}$ defined in Fourier space (see Fig. 2, right).

3. Multifractal sources provide relevant motion information

When dealing with large-scale phenomena in the atmosphere (scales of a few tens of km or more), one usually appreciates several elements that configure the dynamics: almost 2D motion, vertical stratification and also a strong influence of Earth's rotation. Indeed, transport in the atmosphere is dominated by advection, and is quasi-horizontal. However, there are a number of different physical mechanisms by which vertical transport may be present [1]. In this context, we will see that
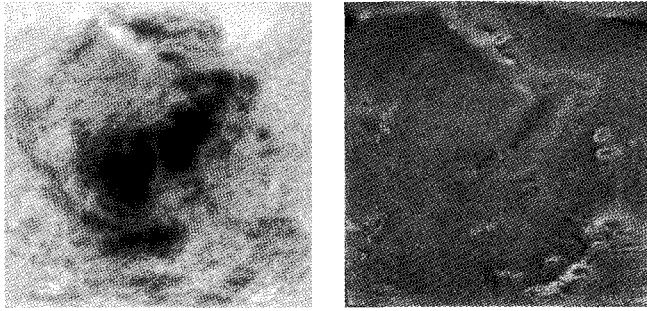
Figure 3.: Left: reduced image computed from the orientated MSM in Fig. 2. Right: representation of the norm of the source field; gray pixels in the sources stand for values close to the average, while zeros are represented by black pixels and poles by white pixels.

the multifractal formalism introduced above can be further completed in order to extract the places where the flow is not contained in the surface of the image.

Indeed, it is possible to obtain more information from the MSM than just reconstructing the signal from it. The detection of the MSM constitutes the stumbling block to infer the properties of the local motion in the flow. It was shown [23] that detecting the MSM is equivalent to detect the main instantaneous (2D planar projection of the) streamlines of the flow. The underlying hypothesis is that vertical movements (mainly inside clouds) modify the spatial variability of temperature [10]. More precisely, as the MSM consists of flow lines, the situation where the flow is purely advective correspond to the situation where the image gradient is perpendicular to the MSM lines with constant modulus. Namely, when temperature is advected by the flow, it is diffused horizontally, i.e. inside planes parallel to the image surface and it is possible to track its movement; however, when temperature is convected, there is a temperature flow transversal, at an unknown rate, to the plane of image acquisition. Thus, for any scalar quantity advected in a region, its average across that region is conserved; more mathematically, the gradient is perpendicular to the streamlines, and with constant modulus. These observations lead to the concept of a *Reduced image R*: in eq. (4), the actual gradient over the MSM ∇I_{h_∞} is replaced with a vector field consistent with the advection hypothesis *i.e.* perpendicular to the MSM and of constant modulus (and with, by convention, the same orientation as the original gradient) [8,23]. Comparing original and proxy signals, we can deduce the temporal evolution of the flow. A simple way to assess if the advective assumption holds in a given area is to verify if the gradient along the MSM is constant and perpendicular to the MSM, *i.e.* if I and R coincide in this area. Where I differs from R, convective movements (perpendicular to the plane of the image) may be present (see Fig. 4).

In order to propose a more quantitative criterion, we use the formalism of sources introduced previously [23]. Let first define the vectorial measure $\vec{\mu}_I$ associated to the signal I as follows. For any subset (region in the image) A, the vectorial
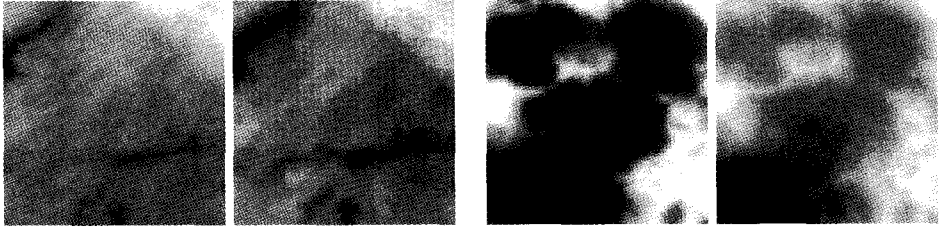
Figure 4.: Comparison of original IR image and its reduced counterpart on two excerpts extracted from Fig. 1 and 3 *resp.* Left: the flow is dominated by advection, the images are rather similar Right: convective activity inside the CS is present, the images are very different. The range of graylevels was expanded in order to show as many details as possible.

measure $\vec{\mu}_I(A)$ is given by:

$$\vec{\mu}_I(A) \equiv \int_A d\vec{y}\, \nabla I(\vec{y}) \tag{5}$$

and an analogous definition for the measure $\vec{\mu}_R$ associated to the reduced signal R. The absolute variations [19] of both $\vec{\mu}_I$ and $\vec{\mu}_R$ are multifractal (in the sense of eq. (2)) with the same singularity exponents $h(\vec{x})$, that is [c]:

$$|\vec{\mu}_I|(B_r(\vec{x})) \sim |\vec{\mu}_R|(B_r(\vec{x})) \sim r^{h(\vec{x})} . \tag{6}$$

Consequently the two measures are absolutely continuous with respect to each other [19], and for that reason one measure can be represented as the integral with the other measure of an appropriate vectorial density $\vec{S}$. We define the vectorial *field of Sources* as the Radon-Nykodin derivative [19]:

$$\vec{S}(\vec{x}) \equiv \lim_{r \to 0} \frac{\vec{\mu}_I(B_r(\vec{x}))}{\vec{\mu}_R(B_r(\vec{x}))} \tag{7}$$

where the limit of the ratio is taken at decreasing scales r. Technically, we make use of the commutative algebra of complex numbers to compute $\vec{S}$: the 2D vector fields defined through eq. (5) are seen as complex fields and the vectorial division in eq. (7) is assumed in the complex plane. The sources are not of multifractal character, as the singularities of I have been removed by dividing by those of R; as a consequence, they are displaying a completely different, more regular structure (Fig. 3, right). It was shown [23] that they enable to easily characterize advection and convection, avoiding the need of considering regions separately and without having to process a sequence. Over advective-dominated areas $\vec{S}$ will be a constant, as the image and its reduced counterpart coincide. On the contrary, on areas dominated by vertical transport, the variations of $\vec{S}$ will be large. Taking into account that roughly $\nabla I \approx \vec{S}\, \nabla R$ (the sources represent the matrix field of transformation between the gradients of I and R), advection means that $\vec{S}$ is a constant, real

---

[c]In theory, R attains the most uniform intensity distribution compatible with the multifractal structure of I.

number. For convection, on the contrary, the highest discrepancy between ∇I and ∇R will happen when one of them vanishes while the other is still finite: hence, the strongest convection appears in the zeroes and poles of $\vec{S}$.

4. Validation and comparison with temporal information

In order to validate the role of the sources, we present evidence on the connection of sources with the dynamics of the atmospheric flow. For that purpose, we compare the features extracted through the estimation of the source field with the description of motion obtained through velocity fields.

Fig. 5 shows two IR samples belonging to a common sequence of images displaying cloudy systems that are known to involve convective activity, namely CS [1]. Hence, some temporal information can be inferred. The most common method for meteorological sequence analysis consists in estimating the apparent motion through dense velocity fields computation. In particular, using optical flow (OF) methodologies [4], it is possible to assign a velocity vector to each point in the image using conservation hypothesis and regularisation constraints. Even if these hypotheses are seldom satisfied in the case of natural fluid flows, OF may provide a reasonable approximation of apparent motion. If we compute the divergence of the resulting flow, we observe (see Fig. 5) a very good correspondence of large divergence values with singular focii in the vector field of sources. The sources correspond to the purely convective area of CS, dominated by vertical motions [23]: singularities of the motion field (that can be connected to convection) are related to the spatial distribution of flow transport across scales. This result is quite promising: sources are computed using one image only, contrary to OF that requires sequence of images, and provide however relevant, comparable information about the temporal evolution of CS.

5. Conclusion

In this paper, we have developed and validated a multifractal technique to detect and locate critical areas of meteorological structures, associated with 2D divergent motion or, equivalently, 3D vertical transport, that are displayed in satellite IR images. Due to the fact that the evolution of atmosphere involves fluid phenomena, the multifractal formalism, derived from thermodynamical concepts, appears as the most suitable approach to analyze such data.

The main result of this study regards the characterization of the, so-called, source field and its specificity. Sources reveal the gradual deviation from advection to convection because the multifractal structure is completely determined by the properties of the MSM. Once we know how to translate any constraint on the dynamics (as for instance advection) to the MSM, we can construct a reduced signal verifying that constraint, and then compare it with the original signal. The Radon-Nikodym derivative will remove the part that both signals share (multifractal structure), highliting their differences (dynamics). The multifractal scheme
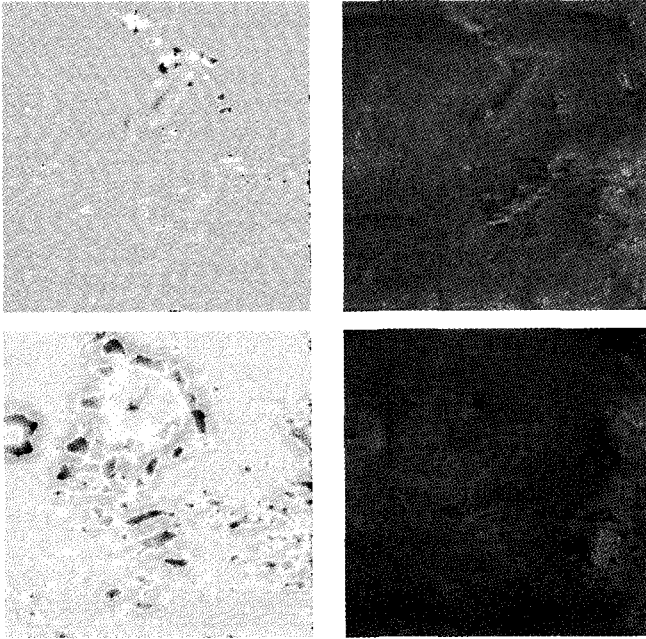
Figure 5.: Comparison of the information provided by velocity field estimation and source field computation on IR images. Selected samples were extracted from a common sequence displaying the temporal evolution of the CS of Fig.1. Left: divergence of the velocity fields estimated with OF approach; non null divergent areas are represented in black (positive) and white (negative); average gray values correspond to areas with (close to) null divergence. Right: norm of the sources.

proposed does not require of any continuity or smoothness hypothesis (as OF methods, for instance) and offers an instantaneous information derived from still images (in opposition to OF, that require a sequence of images sampled at a rate fast enough). The power of multifractal analysis lies in the existence of a hierarchy which is conserved by the dynamics and which can then be used to know more about it. Besides, any other problem in which a multifractal structure has been reported is a good candidate for a multifractal analysis of this kind. Sources can be calculated in order to characterize the underlying dynamical properties. More generally, they provide a new way to analyze and interpret many natural systems.

References

1. Y. Arnaud, M. Desbois, and J. Maizi. Automatic tracking and characterization of african convective systems on MeteoSat pictures. *J. Appl. Meteor.*, 31:443–493, 1992.
2. A. Arneodo, F. Argoul, E. Bacry, J. Elezgaray, and J.F. Muzy. *Ondelettes, Multifractales et Turbulence.* Diderot Editeur, Paris, France, 1995.
3. J. Arrault, A. Arneodo, A. Davis, and A. Marshak. Wavelet-based multifractal analysis of rough surfaces: Application to cloud models and satellite data. *Phys. Rev. Lett.*, 79:75–79, 1997.

4. J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *Int. J. Comp. Vis.*, 12:43–77, 1994.

5. L.M. Carvalho, D. Lavallée, and C. Jones. Multifractal properties of evolving convective systems over tropical South America. *Geophys. Res. Lett.*, 29(10), 2002.

6. A. Davis, A. Marshak, W.J. Wiscombe, and R.F. Cahalan. Multifractal characterizations of nonstationarity and intermittency in geophysical fields, observed, retrieved or simulated. *J. Geophys. Res.*, 99:8055–8072, 1994.

7. U. Frisch. *Turbulence. The Legacy of Kolmogorov.* Cambridge University Press, Cambridge MA, 1995.

8. J. Grazzini, A. Turiel, and H. Yahia. Entropy estimation and multiscale processing in meteorological satellite images. In *Proc. ICPR*, volume 3, pages 764–768, 2002.

9. V.-K. Gupta and E.-C. Waymire. A statistical analysis of mesoscale rainfall as a random cascade. *J. Appl. Meteor.*, 32:251–267, 1993.

10. R.A. Houze. *Clouds Dynamics.* Academic Press, New-York, 1993.

11. D.S. Lavallée, S. Lovejoy, D. Schertzer, and P. Ladoy. Nonlinear variability of landscape topography: Multifractal analysis and simulation. In L. De Cola and N. Lam, editors, *Fractals in Geography*, pages 158–192, New-York, 1993. Prentice Hall.

12. D.C. Lin and R.L. Hughson. Extracting multiple scaling in long-term heart rate variability. In Novak [17].

13. S. Lovejoy and D. Schertzer. Multifractal, universality classes and satellite and radar measurement for rain and cloud fields. *J. Geophys. Res.*, 95:2021–2034, 1990.

14. S.G. Mallat. A theory of multiresolution signal decomposition. *IEEE Trans. on PAMI*, 11(7):674–693, 1989.

15. B. Mandelbrot. *Les Objets Fractals. Forme, Hasard et Dimension.* Nouvelle Bibliothque Scientifique, Flammarion, Paris, 1975.

16. B. Mandelbrot. *Fractals and Scaling In Finance.* Springer-Verlag, New York, 1997.

17. M. Novak, editor. *Paradigms of Complexity. Fractals and Structures in the Sciences,* Singapore, 2000. World Scientific.

18. G. Parisi and U. Frisch. On the singularity structure of fully developed turbulence. In M. Ghil, R. Benzi, and G. Parisi, editors, *Turbulence and Predictability in Geophysical Fluid Dynamics.*, pages 84–87, Amsterdam, 1985. North Holland.

19. W. Rudin. *Real and Complex Analysis.* Mc Graw Hill, New York, USA, 1987.

20. D. Schertzer and S. Lovejoy. *Nonlinear Variability in Geophysics - Scaling and Fractals.* Kluwer Academic, Dordrecht, Holland, 1991.

21. H.E. Stanley. Powerlaws and universality. *Nature*, 378:554, 1995.

22. A. Turiel and A. del Pozo. Reconstructing images from their most singular fractal set. *IEEE Trans. on Im. Proc.*, 11:345–350, 2002.

23. A. Turiel, J. Grazzini, and H. Yahia. Multiscale techniques for the detection of precipitation using thermal IR satellite images. To app. in IEEE Lett. Geosc. Rem. Sens.

24. A. Turiel, G. Mato, N. Parga, and J.-P. Nadal. The Self-Similarity properties of natural images resemble those of turbulent flows. *Phys. Rev. Lett.*, 80(5):1098–1101, 1998.

25. A. Turiel and N. Parga. The multi-fractal structure of contrast changes in natural images: From sharp edges to textures. *Neur. Comp.*, 12:763–793, 2000.

THE DISTANCE RATIO FRACTAL IMAGE

XI-ZHE ZHANG AND ZHENG-XUAN WANG

College of Computer Science and Technology, Jilin University,
Changchun, People's Republic of China
E-mail: zxzok@163.com

TIAN-YANG LV

College of Computer Science and Technology,Harbin Engineering University,
Harbin, People's Republic of China
E-mail: raynor1979@163.com

The iteration of complex function can generate beautiful fractal images. This paper extends the existing iteration methods and tries to find a new way to generate fractal image. We present a novel method based on the iteration of the distance ratio with two points, which classifies the points according to their distance ratio iteration convergence speed. By using this method, we get a 4-dimension fractal set. Firstly, this paper proves the condition of distance ratio convergence and gives an inverse iteration layered method to render the distance ratio fractal. Secondly, taking the quadratic complex function for example, we render the distance ratio fractal image by projecting the distance ratio fractal set into 2-dimension plane. Finally, we analyze some properties of distance ratio fractal image and prove its boundary is the Julia set. Comparing with traditional escape time algorithm, the new method uses iteration of two points instead of one point to generate more complex structure, which is different from existing fractal image.

1. Introduction

Since the introduction of fractal by Mandelbrot [1], fractals have experienced considerable success in quantifying the complex structure exhibited by many natural patterns and have captured the imaginations of both scientists and artists. Fractal art is an algorithmic approach for producing computer-generated art. Many researchers perform research on the generation method of fractal [2] and draw many beautiful images [3].

The iteration of complex function can generate beautiful and complex fractal images, such as M-J sets using escape time algorithm. However, it is difficult to render the image of convergence region using escape time algorithm due to its one point iteration. The iteration points converge or diverge very fast, so it is hard to get their convergence property and to draw the image in the convergence region. Hooper[4] adds some function transform to escape time algorithm to draw internal structures, but it cannot distinguish the convergence speed of the iteration point either.

This paper presents a novel method based on the iteration of distance ratio with two points, which can generate a new type fractal, named as distance ratio fractal. We research the iteration property of distance ratio and prove its convergence condition. Based on the convergence property of distance ratio, we generate a fractal set with its convergence time. Using the inverse iteration layered method, a distance ratio image can be rendered. As this image has unlimited details and self-similarity structure, it is a fractal image. We analyze some properties of distance ratio fractal image and prove its boundary is the Julia set.

The paper is organized as follows: Section 2 introduces some mathematical notation; Section 3 states the definition of distance ratio and proves its convergence condition; Section 4 takes the complex mapping $f(z) = z^2 + c$ as an example and states the algorithm for rendering the distance ratio image; Section 5 proves some properties of distance ratio fractal image; Finally, section 6 summarizes the paper.

2. Basic Concepts

This section reviews some of the basic mathematical notation that will be used throughout the paper.

In mathematics, a contraction mapping on a metric space (X, d) is a function f from X to itself, with the property that there is some real number $s < 1$ that for all x and y in X, $d(f(x), f(y)) \leq sd(x, y)$. The number s is termed as contractivity factor.

An important property of contraction mappings is given by the Banach fixed-point theorem. The theorem guarantees the existence and uniqueness of fixed-points of contraction mappings of metric spaces, and provides a constructive method to find those fixed points.

Let (X, d) be a non-empty complete metric space. Let $f : X \to X$ be a contraction mapping on X. Then the map f admits one and only one fixed-point x^* in X which means $f(x^*) = x^*$. Furthermore, this fixed-point can be found as follows: start with an arbitrary element x_0 in X and define a sequence by $x_n = f(x_{n-1})$ for $n = 1, 2, 3...$ This sequence converges, and its limit is x^*. The following inequality describes the speed of convergence:

$$d(x^*, x_n) \leq \frac{k^n}{1-k} d(x_1, x_0)$$

In this paper we discuss the contraction mapping on C (space of complex numbers).

Julia sets are fractal shapes defined on C. Given an iterated map of the complex plane to itself, the Julia set can be defined as follow [2]:

Let $f^{(p)}$ be the k-fold of function f. If $f^{(p)}(z) = z$ for some integer $p \geq 1$, we call z a periodic point of f, the least p satisfies $f^{(p)}(z) = z$ is called the period of f. Let z be a periodic point of f, with $\left(f^{(p)}\right)'(z) = \lambda$. If $|\lambda| > 1$, call z is repelling. The Julia set $J(f)$ may be defined as the closure of the set of repelling periodic points of f.

Let z^* be the fixed-point of function f, define $A(z^*) = \{z \in C : f^{(k)}(z) \to z^*, (k \to \infty)\}$ as the basin of attraction of fixed point z^*.

Suppose z^* is the fixed-point of f, then the attraction basin of z^* is $J(f)$.

3. Distance Ratio and its Iteration

Let $f : C \to C$ be an analytic function, for any $z_1 \neq z_2 \in C$, define the distance ratio $L(z_1, z_2)$ as follow:

$$L(z_1, z_2) = \frac{|f(z_1) - f(z_2)|}{|z_1 - z_2|} \tag{1}$$

Perform the iteration on the initial points z_1 and z_2, let $z_1^{(n+1)} = f(z_1^{(n)})$, $z_2^{(n+1)} = f(z_2^{(n)}), n = 0, 1, \cdots$, and define the distance ratio after n-times iteration $L^{(n)}(z_1, z_2)$ as:

$$L^{(n)}(z_1, z_2) = \frac{\left| z_1^{(n+1)} - z_2^{(n+1)} \right|}{\left| z_1^{(n)} - z_2^{(n)} \right|} \tag{2}$$

$L(z_1, z_2)$ represents the contractive rate of initial points distance after iteration. According to the definition of derivative, when z_1 approaches z_2, $L(z_1, z_2)$ approaches the derivative of $f(z_2)$. Changing the way of approach into iteration, the initial points will approach the fixed-point if f is a contraction mapping; $L^{(n)}(z_1, z_2)$ will converge to the derivative at the fixed-point if n approaches infinity. It will be proved by following theorem:

Theorem 3.1. *If function f has a fixed-point z^* and $A(z^*)$ is the attraction basin of z^*, for any z_1 and z_2 in $A(z^*)$, it follow*

$$\lim_{k \to \infty} L^{(k)}(z_1, z_2) = |f'(z^*)| \tag{3}$$

Proof: For z_1 and z_2 are the points in $A(z^*)$, we have $\lim_{k \to \infty} z_1^{(k)} \to z^*$ and $\lim_{k \to \infty} z_2^{(k)} \to z^*$. Let $\Delta z = z_1^{(k)} - z_2^{(k)} \to 0$, compute the iteration of distance ratio:

$$\lim_{k \to \infty} L^{(k)}(z_1, z_2) = \frac{\left| \lim_{k \to \infty} z_1^{(k+1)} - \lim_{k \to \infty} z_2^{(k+1)} \right|}{\left| \lim_{k \to \infty} z_1^{(k)} - \lim_{k \to \infty} z_2^{(k)} \right|} = \left| \lim_{\Delta z \to 0} \frac{f(z^* + \Delta z) - z^*}{\Delta z} \right| = |f'(z^*)|$$

The iteration times that $L(z_1, z_2)$ approaches the derivative, are different for different initial points. Thus, the approaching times are related to initial points and the following theorem is stated:

Theorem 3.2. *For a specified threshold ε, $\exists k^* \in N$, if $k > k^*$, $\left| L^{(k)}(z_1, z_2) - |f'(z^*)| \right| < \varepsilon$*

Proof: For any $z \in C$, the iteration sequence $z^{(n)} = f(z^{(n-1)})$, $n = 0, 1, \cdots$ is obtained. Let s be the contractivity factor.

$$\left| z^{(n+1)} - z^{(n)} \right| = \left| f\left(z^{(n)}\right) - f\left(z^{(n-1)}\right) \right| \leq s \left| z^{(n)} - z^{(n-1)} \right| \leq \cdots \leq s^n \left| z^{(1)} - z \right|$$

$$\left| z^{(m)} - z^{(n)} \right| \leq \sum_{i=n}^{m-1} \left| z^{(i+1)} - z^{(i)} \right| \leq \sum_{i=n}^{m-1} s^i \left| z^{(1)} - z \right| \leq \frac{s^n}{1-s} \left| z^{(1)} - z \right|, (m > n)$$

Let $m \to \infty$, then

$$\left| z^{(n)} - z^* \right| \leq \frac{s^n}{1-s} \left| z^{(1)} - z \right|$$

Let $\left| z^{(n)} - z^* \right| \leq \varepsilon$, the iteration times can be decided according to

$$s^n \leq \frac{\varepsilon(1-s)}{|z - f(z)|}$$

Suppose the iteration times are k_1 for z_1 and k_2 for z_2, the threshold are ε_1 for z_1 and ε_2 for z_2, let $k > max(k_1, k_2)$, $z_1^{(k)} = z^* + \varepsilon_1^{(k)}, z_2^{(k)} = z^* + \varepsilon_2^{(k)}, |\varepsilon| \leq |\varepsilon_1^{(k)} - \varepsilon_2^{(k)}|$, it follows:

$$\left| L^{(k)}(z_1, z_2) - |f'(z^*)| \right| = \left| \left| \frac{f\left(z^* + \varepsilon_1^{(k)}\right) - f\left(z^* + \varepsilon_2^{(k)}\right)}{\varepsilon_1^{(k)} - \varepsilon_2^{(k)}} \right| - \left| \frac{f(z^* + \varepsilon) - f(z^*)}{\varepsilon} \right| \right|$$

$$\leq \left| \frac{f\left(z^* + \varepsilon_1^{(k)}\right) - f\left(z^* + \varepsilon_2^{(k)}\right) - f(z^* + \varepsilon) + f(z^*)}{\varepsilon_1^{(k)} - \varepsilon_2^{(k)}} \right|$$

$$\leq \left| \frac{s\left(\varepsilon_1^{(k)} - \varepsilon^{(k)} - \varepsilon\right)}{\varepsilon_1^{(k)} - \varepsilon_2^{(k)}} \right| \leq \left| 1 - \frac{\varepsilon}{s^{k-1}(z_1 - z_2)} \right| \leq \varepsilon$$

And the iteration times can be estimated through

$$s^{k-1} \leq \frac{\varepsilon(z_1 - z_2)}{1 + \varepsilon}$$

Define k^* in Theorem 3.2 as the stabilizing times. For a specified threshold ε, $L(z_1, z_2)$ can approach to the derivative in limited times. Obviously, stabilizing times are related to the choice of initial points and are different for the points located in different regions.

4. Distance Ratio Fractal Image of Quadratic Map

Taking the quadratic function $f(z) = z^2 + c$ as an example, we perform researches on the relationship between the stabilizing time k^* and the initial points by iterating the distance ratio. It shows that the method results in a fractal set. And we will analyze its properties.

4.1. *Distance Ratio of Mapping* $f(z) = z^2 + c$

To compute the distance ratio of $f(z) = z^2 + c$, we choose any two points $z_1(x_1, y_1)$ and $z_2(x_2, y_2)$ with $z_1 \neq z_2$, then compute their distance after one time iteration.

$$
\begin{aligned}
|f(z_1) - f(z_2)| &= \sqrt{(x_1^2 - y_1^2 - x_2^2 + y_2^2) + (2x_1y_1 - 2x_2y_2)^2} \\
&= \sqrt{\left[(x_1 - x_2)^2 + (y_1 - y_2)^2\right]\left[(x_1 + x_2)^2 + (y_1 + y_2)^2\right]} \\
&= |z_1 - z_2|\,|z_1 + z_2|
\end{aligned}
$$

Therefore, the distance ratio $L(z_1, z_2) = |z_1 + z_2|$. Perform the iterations on the distance ratio, we have:

$$
\lim_{k \to \infty} L^{(k)}(z_1, z_2) = \left| \lim_{k \to \infty} z_1^{(k)} + \lim_{k \to \infty} z_2^{(k)} \right| = 2|z^*| = |f'(z^*)|
$$

Figure 3 shows the fluctuation of distance ratio during iterations. We can see $L(z_1, z_2)$ approaches the derivative at the fixed-point after seven iterations.
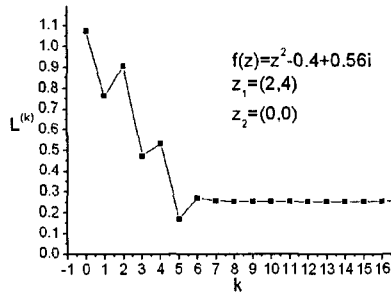
Figure 1.   The fluctuation of distance ratio in iterating

4.2. *Distance Ratio Zone*

Based on the discussion of Section 3, the stabilizing time k^* is different for different initial points. We put the points with the same stabilizing time to one set, named as $zone(k)$. Therefore, the set of all convergence points consists of $zone(0), zone(1), \cdots, zone(\infty)$. We name the set generated this way as the Distance Ratio Set(DRS).

According to Eq.1, the distance ratio is computed by two points $z_1(x_1, y_1)$ and $z_2(x_2, y_2)$. So we have a 4-D set (x_1, y_1, x_2, y_2). In order to render fractal image in 2-D plane, we have one point fixed as the fixed-point and let the other point change. The definition of $zone(k)$ and DRS is as follows:

Supposing z is a complex number and ε is a small real number, we say that z is k-level if z meets the following condition:

$$\left|L^{(k-1)}(z, z^*) - |f'(z^*)|\right| > \varepsilon, \left|L^{(k)}(z, z^*) - |f'(z^*)|\right| \leq \varepsilon \tag{4}$$

The set consisting of all the k-level points is termed $zone(k)$. The set of all the converged points is called the distance ratio set (DRS), which can be divided into some $zone(k)$, like:

$$DRS(f) = \bigcup_{k=0}^{n} zone(k) \tag{5}$$

From above definition we can see that the points in $zone(k)$ approach the derivative after at least k iterations. The relationship of the points in the neighbor zone is stated by the following theorem:

Theorem 4.1. *If f has inverse mapping, for any z in $zone(k+1)$, there exists z' in $zone(k)$, which satisfies $f^{-1}(z') = z$.*

Proof: For any $z \in zone(k+1)$ according to the Eq.1 ,we have,

$$\left|L^{(k)}(z, z^*) - |f'(z^*)|\right| = \left|L^{(k)}(f^{-1}(z'), z^*) - |f'(z^*)|\right| = \left|L^{(k-1)}(z, z^*) - |f'(z^*)|\right| > \varepsilon$$

$$\left|L^{(k+1)}(z, z^*) - |f'(z^*)|\right| = \left|L^{(k+1)}(f^{-1}(z'), z^*) - |f'(z^*)|\right| = \left|L^{(k)}(z, z^*) - |f'(z^*)|\right| \leq \varepsilon$$

So z' satisfies the definition of $zone(k)$, $z' \in zone(k)$. For any z in $zone(k)$, $f(z) \in zone(k-1)$, therefore $zone(k)$ turns into $zone(k-1)$ in one iteration, viz. $f(zone(k)) = zone(k-1)$.

From Theorem 4.1, performing inverse iteration on $zone(k)$ will generate $zone(k+1)$. Starting with $zone(0)$ and repeating this process, we can compute $zone(1), zone(2), \cdots, zone(n)$. Combining the entire zone, we will get the whole distance ratio set.

4.3. Rendering Method

We may have a very complex image by drawing different colors in different zones of the distance ratio set. We name this image as the distance ratio fractal image(DRF). This section will introduce the rendering method of DRF. Firstly, we discuss how to compute $zone(0)$.

Based on Eq.1 and Eq.3, we have:

$$\left|\frac{|f(z) - f(z^*)|}{|z - z^*|} - |f'(z^*)|\right| \leq \varepsilon$$

Let $f(z) = z^2 + c$ and solve above inequation, it follow:

$$2|z^*| - \varepsilon \leq |z + z^*| \leq 2|z^*| + \varepsilon$$

So $zone(0)$ is a ring between two circles centered at $-z^*$, and the radius of one circle is $2|z^*| - \varepsilon$, the other is $2|z^*| + \varepsilon$.

We compute $zone(0)$ firstly, then get $zone(1), zone(2) \cdots$ by performing inverse iteration, and so on to get the whole distance iterated image. We call this method Inverse Iteration Layer Method (IILM). The specific steps of the algorithm IILM are listed as follows:

(1) Specify the max iteration times n and let $k = 0$;
(2) Compute the $zone(0)$;
(3) Perform inverse iteration for all points in $zone(k)$, which will result in the $zone(k+1)$;
(4) Draw $zone(k)$ with different colors that can be got by k;
(5) If $k < n$, then $k = k + 1$, go to (3);

Figure 2 shows the rendering process of DRF. The whole process goes from inner layer to outer layer, like blooming flower. It is also feasible to choose any of the layers to render.

Figure 3 shows some DRF with typical parameter , $\varepsilon = 0.0000001$ and $n = 50$. Fig 3.a shows the image with very small $|c|$. The image is approximate to a circle; many "buds" sequences are located around the center. The number of "buds" of each layer increases greatly and approaches infinity. In Fig 3.b, parameter $|c|$ increases moderately and the structure of the image is diverse. The "buds" structure partly changes to the "hair" structure, and these two structures are co-existing. The border of the image is a simple curve. The value of $|c|$ is big in Fig 3.c, which results in huge change of the image's border and the "buds" structure fully turns into the "hair" structure.

It can be seen from above images that the DRF have following features:

There are little changes in the center area, while the surrounding areas have clear detailed structure, and the further area is located from center, the more complex the mini structure is; "buds" structures exist in the image and its number increases as its distance from center increases. With the increase of c, the "buds" structure turns into the "hair" structure; if c is big enough, these "hair" structures will twist together.

5. The Properties of Distance Ratio Fractal

In this section, we will prove the image rendering in above sections is fractal image. Firstly, we prove the border of the DRF is a Julia set generated by mapping f.

For any point z in DRF, based on Theorem 3.1, if the iteration of distance ratio converges to the derivative at fixed-point, the initial point must converge to fixed-point, viz.

$$\lim_{k \to \infty} L^{(k)}(z_1, z_2) = |f'(z^*)| \Leftrightarrow \lim_{k \to \infty} f^{(k)}(z) = z^*$$

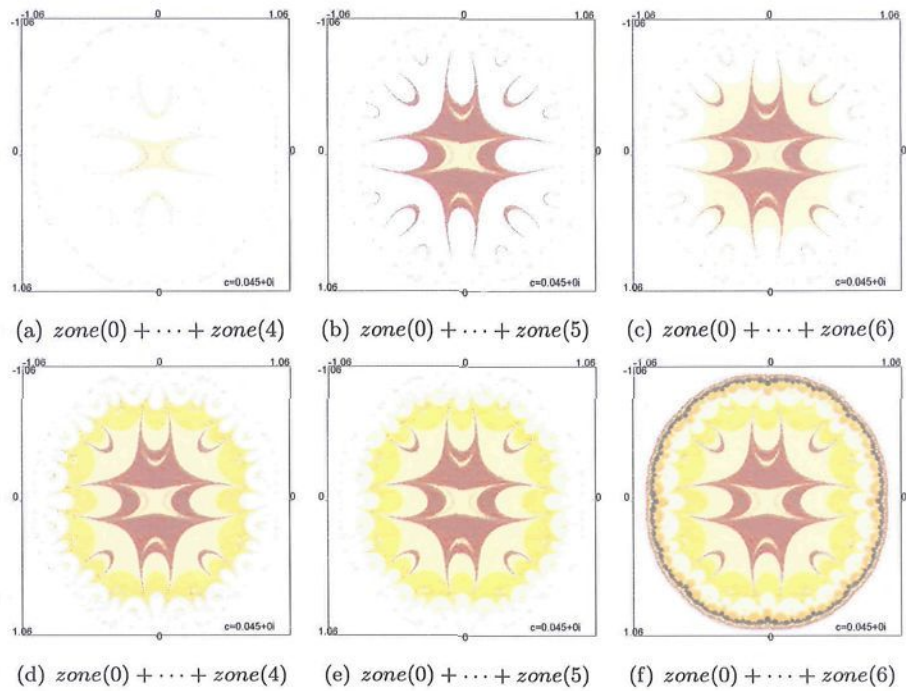(a) $zone(0) + \cdots + zone(4)$ (b) $zone(0) + \cdots + zone(5)$ (c) $zone(0) + \cdots + zone(6)$

(d) $zone(0) + \cdots + zone(4)$ (e) $zone(0) + \cdots + zone(5)$ (f) $zone(0) + \cdots + zone(6)$

Figure 2.   The rendering step of DRF

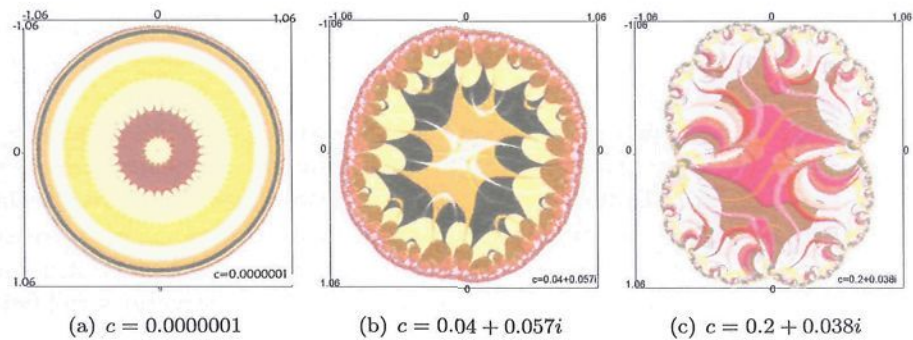(a) $c = 0.0000001$ (b) $c = 0.04 + 0.057i$ (c) $c = 0.2 + 0.038i$

Figure 3.   DRF with difference parameter

So z is located in the attraction basin of z^* and DRF is the attraction basin of the fixed-point.

The border of the attractive basin for the fixed-point of f is the Julia set formed by mapping f^2, thus the border of the DRF is a Julia set of mapping f. Therefore the algorithm of rendering DRF can be adopted as a new method to draw the inner structure of Julia sets. And many conclusions of Julia sets are adaptable. For instance, there is an attractive fixed-point, if c is located in the heart-shape line $z = e^{i\theta} \left(1 - e^{i\theta}/2\right)/2, (0 \le \theta \le 2\pi)^2$, the DRF can be created.

In the following, we state DRF has unlimited details and self-similarity structure. Based on Theorem 3.2, the smaller the threshold ε is, the bigger the stabilizing times k is; when $\varepsilon \to 0$, $k \to \infty$. As $f^{-1}(z) = \pm\sqrt{z-c}$ is multi-value mapping, during process of inverse iteration, the number of points in $zone(k)$ increases doubly with the increase of k. According to above, DRF has unlimited details. Based on Theorem 4.1 we have $f(zone(k+1)) = zone(k)$, so DRF has self-similarity structure. Therefore, it is a fractal image.

The method to render DRF can be extended to many mappings. From above discussions, the distance ratio will converge when mapping has an attractive fixed-point; the detailed structure of each layer will increase doubly if the inverse function of the mapping is multi-valued. Based on these observations, we make the supposition that the mapping satisfying above conditions can create distance ratio fractal image. Figure 4 is an iteration image of another polynomial function . There will be further study of its properties in future works.
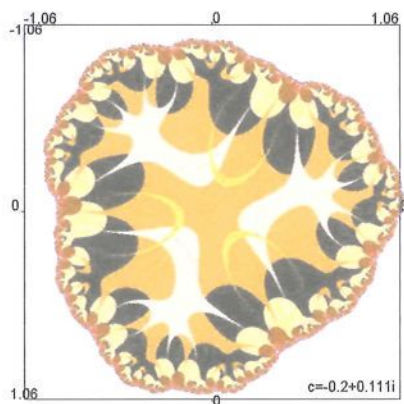
Figure 4. The distance ratio image of $g(z) = z^3 + 0.2 + 0.1i$

6. Conclusion and Future Works

This paper investigates the distance ratio and its properties. If the complex function is contraction mapping, the iteration of distance ratio will converge to the derivative at the fixed-point. By rendering the initial points with different color according to their iteration convergence times, a fractal image is obtained. The image has unlimited detail and self-similarity. Primary analysis shows its properties and proves that its border is the Julia set. The paper researches the properties of DRF in 2-D plane; Future works will do the following:

(1) render the 3-D DRF and analysis their properties;
(2) investigate the properties of DRF generated by other functions;
(3) perform further study of the concrete structure of the fractal images.

References

1. Mandelbrot, B.B, *The fractal geometry of nature*, New York Freeman (1977).
2. Kenneth Falconer, *Fractal Geometry - Mathematical Foundations and Applications*, John Wiley & Sons (1990).
3. *http://www.fractal-art.com/*
4. Kenneth J. Hooper, *A note on some internal structures of the Mandelbrot Set*, Computers & Graphics, **15(2)** (1991)

ITERATED FUNCTION SYSTEMS IN MIXED EUCLIDEAN AND 𝔭-ADIC SPACES

BERND SING

Fakultät für Mathematik
Universität Bielefeld
Universitätsstraße 25
33615 Bielefeld
Germany
E-mail: sing@math.uni-bielefeld.de
URL: http://www.mathematik.uni-bielefeld.de/baake/sing/

We investigate graph-directed iterated function systems in mixed Euclidean and 𝔭-adic spaces. Hausdorff measure and Hausdorff dimension in such spaces are defined, and an upper bound for the Hausdorff dimension is obtained. The relation between the Haar measure and the Hausdorff measure is clarified. Finally, we discus an example in $\mathbb{R} \times \mathbb{Q}_2$ and calculate upper and lower bounds for its Hausdorff dimension.

1. Introduction and Setting

The main focus of this article is the following situation: Assume that a (finite) family $(\Omega_1, \ldots, \Omega_n)$ of subsets of a locally compact Abelian group $\mathbb{X}$, the topology of which is assumed to be generated by a metric, is implicitly given as the unique solution of a graph-directed iterated function system (GIFS). Can we define and calculate the Hausdorff measure and Hausdorff dimension of these sets, and determine their relation to the Haar measure in $\mathbb{X}$?

In the following, we assume that the space $\mathbb{X}$ is given by

$$\mathbb{X} = \mathbb{R}^r \times \mathbb{C}^s \times \mathbb{Q}_{\mathfrak{p}_1} \times \cdots \times \mathbb{Q}_{\mathfrak{p}_k}, \tag{1}$$

i.e., as a product of non-discrete locally compact fields (we shall expand on $\mathbb{Q}_{\mathfrak{p}}$ below). We call the number

$$\dim_{\mathrm{metr}} \mathbb{X} = r + 2 \cdot s + k \tag{2}$$

the *metric dimension* of $\mathbb{X}$ (also see Section 3).

The organisation of this article is as follows: To keep everything as self-contained as possible, we briefly review 𝔭-adic spaces in Section 2. In Section 3, the relation

267

between Hausdorff measure and Haar measure on $\mathbb{X}$ is clarified. Iterated function systems on $\mathbb{X}$ are introduced in Section 4. We define the affinity dimension for a GIFS and show that it is an upper bound for the Hausdorff dimension of the sets Ω_i. We also discus a condition for which we obtain a lower bound for the Hausdorff dimension. In the last Section, we explore a GIFS in $\mathbb{R} \times \mathbb{Q}_2$.

2. p-adic Spaces and their Visualisation

An *algebraic number field* K is a finite field extension of $\mathbb{Q}$ lying in $\mathbb{C}$, i.e., it is a simple extension of the form $K = \mathbb{Q}(\lambda)$. The *integral closure* of $\mathbb{Z}$ in an algebraic number field K is called the ring of *algebraic integers* $\mathfrak{o}_K$ of K. An ideal $\mathfrak{p}$ of the ring $\mathfrak{o}_K$ is called *prime* if the quotient $\mathfrak{o}_K/\mathfrak{p}$ is an integral domain. A key theorem[10] in algebraic number theory states that every (fractional) ideal of $\mathfrak{o}_K$ in K can be uniquely factored into prime ideals.

Let K^* be the multiplicative group of non-zero elements of K. A surjective homomorphism $v : K^* \to \mathbb{Z}$ with $v(x + y) \geq \min\{v(x), v(y)\}$ (and the convention $v(0) = \infty$) is called a *valuation*[18]. Every prime ideal $\mathfrak{p}$ yields a valuation of K, called the p-adic valuation $v_\mathfrak{p}$, and these are all possible valuations: For $x \in K$ let $v_\mathfrak{p}(x) = v_\mathfrak{p}(x\mathfrak{o}_K)$ (i.e., $x\mathfrak{o}_K$ is the (fractional) ideal generated by x) where a (fractional) ideal $\mathfrak{a}$ has the unique factorisation $\mathfrak{a} = \mathfrak{p}_1^{v_{\mathfrak{p}_1}(\mathfrak{a})} \cdots \mathfrak{p}_\ell^{v_{\mathfrak{p}_\ell}(\mathfrak{a})}$ into prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_\ell$.

Given a p-adic valuation $v_\mathfrak{p}$, one obtains an *ultrametric absolute value* (or, more precisely, a *non-Archimedean absolute value*) by $\|x\|_\mathfrak{p} = \eta^{-v_\mathfrak{p}(x)}$ for some $\eta > 1$ (where $\|0\|_\mathfrak{p} = 0$). The completion of $K = \mathbb{Q}(\lambda)$ with respect to such a p-adic absolute value yields the p-*adic number field* $\mathbb{Q}_\mathfrak{p}$, which is a locally compact field. We note that the completion of $\mathbb{Q}$ itself w.r.t. the prime ideal $p\mathbb{Z}$ yields the p-adic numbers $\mathbb{Q}_p$.

We define the p-adic integers $\mathbb{Z}_\mathfrak{p} = \{x \in \mathbb{Q}_\mathfrak{p} \mid \|x\|_\mathfrak{p} \geq 0\}$ and the related ideal $\mathfrak{m}_\mathfrak{p} = \{x \in \mathbb{Q}_\mathfrak{p} \mid \|x\|_\mathfrak{p} > 0\}$. Then $\mathbb{Z}_\mathfrak{p}$ is a *discrete valuation ring*, i.e., it is a principal ideal domain that has a unique non-zero prime ideal, namely $\mathfrak{m}_\mathfrak{p}$. Furthermore, the *residue field* $k_\mathfrak{p} = \mathbb{Z}_\mathfrak{p}/\mathfrak{m}_\mathfrak{p}$ is finite, and the choice $\eta = [\mathbb{Z}_\mathfrak{p} : \mathfrak{m}_\mathfrak{p}]$ in the above definition of the p-adic absolute value yields the so-called *normalised* p-*adic absolute value* (which has nice properties w.r.t. the Haar measure on $\mathbb{Q}_\mathfrak{p}$, see Section 3).

An element π which generates $\mathfrak{m}_\mathfrak{p}$, i.e., $\mathfrak{m}_\mathfrak{p} = \pi\mathbb{Z}_\mathfrak{p}$, is called a *uniformizer*. By the uniqueness of $\mathfrak{m}_\mathfrak{p}$, the non-zero ideals of $\mathbb{Z}_\mathfrak{p}$ are given by $\pi^m\mathbb{Z}_\mathfrak{p}$ ($m \in \mathbb{N}_0$). If S is a system of representatives of $k_\mathfrak{p}$ (including 0 for simplicity), every element $x \in \mathbb{Q}_\mathfrak{p}$ can be written uniquely as a convergent series (w.r.t. the p-adic absolute value)

$$x = \sum_{j=m}^{\infty} s_j \, \pi^j, \tag{3}$$

with $s_j \in S$ and $m \in \mathbb{Z}$. If $x \in \mathbb{Z}_\mathfrak{p}$, then one can take $m = 0$ and we simply write (with obvious meaning) $x = s_0 s_1 s_2 \ldots$.

One can visualise $\mathbb{Z}_\mathfrak{p}$ (and also $\mathbb{Q}_\mathfrak{p}$) as a Cantor set[16]. For example, if we take $\mathbb{Q}_2$, every $x \in \mathbb{Q}_2$ can be written as $x = \sum_{j=m}^{\infty} s_j \, 2^j$, where $S = \{0, 1\}$. Therefore, $\mathbb{Z}_2$ can be identified with the set of all 0-1-sequences, i.e., $\mathbb{Z}_2 = \{0, 1\}^{\mathbb{N}_0}$. But this is also a coding of points in the Cantor set[a], and points which are close in the Cantor set are also close w.r.t. the 2-adic metric (also, both the Cantor set and $\mathbb{Q}_2$ are totally disconnected). In Section 5, we will use the Cantor set to visualise sets in $\mathbb{Z}_2$, where (for reasons of representation) we take a factor of $\frac{1}{2}$ instead of $\frac{1}{3}$ in the construction of the "Cantor set" (of course, one then obtains the whole interval $[0, 1]$).

3. Haar and Hausdorff Measures

Given an Abelian topological group G, a measure μ on the family $\mathcal{B}$ of Borel sets in G is called a *Haar measure* if it satisfies the following conditions[6,12,5]:

H1 μ is a regular measure.
H2 If C is compact, $\mu(C) < \infty$.
H3 μ is not identically zero.
H4 μ is invariant under translations, i.e., $\mu(B + t) = \mu(B)$ for all $B \in \mathcal{B}$ and $t \in G$.

Haar measures are unique up to a multiplicative constant. They are obtained by a so-called "Method I Construction"[12].

The Haar measure on $\mathbb{X}$ is the product measure of the Haar measures of its factors. We remark that the (1-dimensional) Lebesgue measure on $\mathbb{R}$ and the 2-dimensional Lebesgue measure on $\mathbb{R}^2 \simeq \mathbb{C}$ are Haar measures. We also note that we have for a Haar measure μ on $\mathbb{R}$, resp. $\mathbb{C}$, resp. $\mathbb{Q}_\mathfrak{p}$

- $\mu(\alpha B) = |\alpha| \cdot \mu(B)$ if $\alpha \in \mathbb{R}$ and $B \subset \mathbb{R}$.
- $\mu(\alpha B) = |\alpha|^2 \cdot \mu(B)$ if $\alpha \in \mathbb{C}$ and $B \subset \mathbb{C}$.
- $\mu(\alpha B) = \|\alpha\|_\mathfrak{p} \cdot \mu(B)$ if $\alpha \in \mathbb{Q}_\mathfrak{p}$ and $B \subset \mathbb{Q}_\mathfrak{p}$ (where $\| \cdot \|_\mathfrak{p}$ denotes the normalised $\mathfrak{p}$-adic absolute value).

On the other hand, $\mathbb{X}$ is also a separable metric space, where we take the maximum metric d_∞, i.e., for $x, y \in \mathbb{X}$ with

$$x = (x_1, \ldots, x_r, x_{r+1}^{(1)} + i \cdot x_{r+1}^{(2)}, \ldots, x_{r+s}^{(1)} + i \cdot x_{r+s}^{(2)}, x_{r+s+1}, \ldots, x_{r+s+k}) \quad (4)$$

(where $x_1, \ldots, x_r, x_{r+1}^{(1)}, \ldots, x_{r+s}^{(1)}, x_{r+1}^{(2)}, \ldots, x_{r+s}^{(2)} \in \mathbb{R}$, while $x_{r+s+j} \in \mathbb{Q}_{\mathfrak{p}_j}$ for all $1 \leq j \leq k$) we have

$$\begin{aligned} d_\infty(x, y) = \max\{&|x_1 - y_1|, \ldots, |x_r - y_r|, |x_{r+1}^{(1)} - y_{r+1}^{(1)}|, \ldots, |x_{r+s}^{(1)} - y_{r+s}^{(1)}|, \\ &|x_{r+1}^{(2)} - y_{r+1}^{(2)}|, \ldots, |x_{r+s}^{(2)} - y_{r+s}^{(2)}|, \|x_{r+s+1} - y_{r+s+1}\|_{\mathfrak{p}_1}, \ldots, \\ &\|x_{r+s+k} - y_{r+s+k}\|_{\mathfrak{p}_k}\}. \end{aligned} \quad (5)$$

---

[a]Indeed, the Cantor set is given by $\left\{ x = \frac{2}{3} \cdot \sum_{j=0}^{\infty} s_j \left(\frac{1}{3}\right)^j \mid s_j \in \{0, 1\} \right\}$.

Therefore, we can define the *diameter* of a set $B \subset \mathbb{X}$ by $\mathrm{diam}(B) = \sup_{x,y \in B} d_\infty(x, y)$ with the convention $\mathrm{diam}(\varnothing) = 0$. Then, the measure obtained by the so-called "Method II Construction"[12,17] from the set function $\tau(B) = [\mathrm{diam}(B)]^d$ is a measure and called the *d-dimensional Hausdorff measure* $h^{(d)}$.

Generalising Theorem 30 in Ref. 17, we can show that the two measures are related as follows.

Theorem 3.1. *Let the space* $\mathbb{X}$ *be given as in Eq. (1), and let* $d = \dim_{\mathrm{metr}} \mathbb{X}$. *Then, the d-dimensional Hausdorff measure* $h^{(d)}$ *is a Haar measure. Furthermore,* $h^{(d)}$ *equals the Haar measure constructed as product measure where we assign measure* 1 *to the unit interval (in* $\mathbb{R}$*) resp. to* $\mathbb{Z}_p$ *(in* $\mathbb{Q}_p$*).* □

As usual, it is clear that $h^{(d)}(B)$ is non-increasing for a given subset $B \subset \mathbb{X}$ as d increases from 0 to ∞. Furthermore, there is a unique value $\dim_{\mathrm{Hd}} B$, called the *Hausdorff dimension* of B, such that $h^{(d)}(B) = \infty$ if $0 \le d < \dim_{\mathrm{Hd}} B$ and $h^{(d)}(B) = 0$ if $d > \dim_{\mathrm{Hd}} B$.

Note that one can see from this property that Hausdorff dimension is a metric concept rather than a topological one[7] (therefore we have chosen the name *metric dimension*; the (topological) dimension of $\mathbb{X}$ is $r + 2 \cdot s$, because $\mathfrak{p}$-adic spaces $\mathbb{Q}_\mathfrak{p}$ are totally disconnected).

4. Graph-Directed Iterated Function Systems

Let us consider the following subspace $\mathcal{L}$ of linear mappings[b] from $\mathbb{X}$ to $\mathbb{X}$: For each $T \in \mathcal{L}$, there are numbers $a_1, \ldots, a_{r+s+k}$ such that

$$T(x) = T((x_1, \ldots, x_{r+s+k})) = (a_1 \cdot x_1, \ldots, a_{r+s+k} \cdot x_{r+s+k}), \qquad (6)$$

where $x_1, \ldots, x_r, a_1, \ldots, a_r \in \mathbb{R}$, while $x_{r+1}, \ldots, x_{r+s}, a_{r+1}, \ldots, a_{r+s} \in \mathbb{C}$ and $x_{r+s+j}, a_{r+s+j} \in \mathbb{Q}_{\mathfrak{p}_j}$ $(1 \le j \le k)$.

We now look at the family (complex numbers $a_{r+1}, \ldots, a_{r+s}$ taken twice) of the $r + 2 \cdot s + k$ numbers

$$(|a_1|, \ldots, |a_r|, |a_{r+1}|, |a_{r+1}|, |a_{r+2}|, \ldots, |a_{r+s-1}|,$$
$$|a_{r+s}|, |a_{r+s}|, \|a_{r+s+1}\|_{\mathfrak{p}_1}, \ldots, \|a_{r+s+k}\|_{\mathfrak{p}_k}), \qquad (7)$$

called the *singular values* of T. We order them in descending order $\alpha_1 \ge \alpha_2 \ge \ldots \ge \alpha_{r+2s+k}$, where $(\alpha_1, \ldots, \alpha_{r+2s+k})$ is a permutation of $(|a_1|, \ldots, \|a_{r+s+k}\|_{\mathfrak{p}_k})$. We are only interested in maps $T \in \mathcal{L}$ which are contracting $(\alpha_1 < 1)$ and non-singular $(\alpha_{r+2s+k} > 0)$. We denote the subspace of non-singular and contracting maps of $\mathcal{L}$ by $\mathcal{L}'$.

---

[b]One can also consider more general linear mappings[2]; the ones considered here then correspond to the case where the coordinate axes and the principal axes coincide.

The *singular value function* $\Phi^q(T)$ of $T \in \mathcal{L}'$ is defined[2,3] for $q \geq 0$ as follows:

$$\Phi^q(T) = \begin{cases} 1 & \text{if } q = 0 \\ \alpha_1 \cdot \alpha_2 \cdot \ldots \cdot \alpha_{j-1} \cdot \alpha_j^{q-j+1} & \text{if } j - 1 < q \leq j \\ (\alpha_1 \cdot \alpha_2 \cdot \ldots \cdot \alpha_{r+2s+k})^{q/(r+2s+k)} & \text{if } q > r + 2 \cdot s + k \end{cases} \tag{8}$$

Then, $\Phi^q(T)$ is continuous and strictly decreasing in q. Moreover, for fixed q, the singular value function is submultiplicative, i.e., $\Phi^q(T \circ U) \leq \Phi^q(T) \cdot \Phi^q(U)$ for $T, U \in \mathcal{L}'$. Note that we have $\Phi^q(T^n) = [\Phi^q(T)]^n$.

We now look at a *graph-directed iterated function system* (GIFS) ($1 \leq i \leq n$):

$$\Omega_i = \bigcup_{i=1}^{n} \bigcup_{f_{ij}^{(\ell)} \in F_{ij}} f_{ij}^{(\ell)}(\Omega_j), \tag{9}$$

where F_{ij} is a (finite) set of affine contracting mappings, i.e., $f_{ij}^{(\ell)}(x) = T_{f_{ij}^{(\ell)}}(x) + t_{f_{ij}^{(\ell)}}$ with $T_{f_{ij}^{(\ell)}} \in \mathcal{L}'$ and $t_{f_{ij}^{(\ell)}} \in \mathbb{X}$. A GIFS can be visualised by a directed multi-graph $G_{(\Omega_1, \ldots, \Omega_n)}$, where the vertices are the sets Ω_i. If $F_{ij} \neq \varnothing$, we draw $|F_{ij}|$ directed edges from Ω_i to Ω_j, labelling each edge with exactly one of the maps $f_{ij}^{(\ell)}$. We denote by $\boldsymbol{F}$ the matrix[c] $\boldsymbol{F} = (|F_{ij}|)_{1 \leq i,j \leq n}$ (with the convention $|\varnothing| = 0$) and by $\rho(\boldsymbol{F})$ its spectral radius.

We define the *path space* E^∞ as the set of all infinite paths in the graph along directed edges that start at some vertex. Each path (and its starting point) is (uniquely, maybe after renaming) indexed by the sequence of the edges $\omega = (\omega_1 \omega_2 \ldots)$ it runs along. We also define the sets $E^{(0)} = \varnothing$ (paths of length 0), and the set $E_{ij}^{(\ell)}$ of all paths of length ℓ that start at Ω_i and end at Ω_j (then $\omega_1 \in \bigcup_{m=1}^{n} F_{im}$ and $\omega_\ell \in \bigcup_{m=1}^{n} F_{mj}$). We also set $E^{(\ell)} = \bigcup_{i=1}^{n} \bigcup_{j=1}^{n} E_{ij}^{(\ell)}$ (all paths of length ℓ), $E^{\text{fin}} = \bigcup_{\ell \geq 0} E^{(\ell)}$ (all finite paths) and $E^* = E^{\text{fin}} \cup E^\infty$.

For $\omega \in E^{\text{fin}}$ and $\varpi \in E^*$, we denote by $\omega\varpi$ the sequence obtained by concatenation (or juxtaposition) if $\omega\varpi \in E^*$. If ω is a prefix of ϖ, i.e., $\varpi = \omega \ldots$, we write $\omega < \varpi$. By $\omega \wedge \varpi$ we denote the maximal sequence such that both $(\omega \wedge \varpi) < \omega$ and $(\omega \wedge \varpi) < \varpi$. We can topologise E^∞ in a natural way using the ultrametric $d(\omega, \varpi) = \eta^{-|\omega \wedge \varpi|}$ for some $\eta > 1$. Then E^∞ is a compact space and the sets $N(\varpi) = \{\omega \in E^\infty \mid \varpi < \omega\}$ with $\varpi \in E^{\text{fin}}$ form a basis of clopen sets for E^∞.

For $\omega = (\omega_1 \ldots \omega_\ell) \in E^{\text{fin}}$, we define $T_\omega = T_{\omega_1} \circ \ldots \circ T_{\omega_\ell}$ (with $T_\varnothing(x) = x$), i.e., we are only interested in the linear part of each map $\omega_i(x) = T_{\omega_i}(x) + t_{\omega_i}$. By the "Method II Construction" with the set function $\tau^q(N(\omega)) = \Phi^q(T_\omega)$ (with $\tau^q(\varnothing) = 0$), we obtain a measure $\nu^{(q)}$ on E^∞. Then, we can generalise Proposition 4.1 of Ref. 2.

Proposition 4.1. *For a GIFS (with strongly connected directed graph), the following numbers exist and are all equal:*

---

[c]This is also the adjacency matrix of the graph $G_{(\Omega_1, \ldots, \Omega_n)}$.

(1) $\inf\{q \mid \sum_{\omega \in E^{\mathrm{fin}}} \Phi^q(T_\omega) < \infty\} = \sup\{q \mid \sum_{\omega \in E^{\mathrm{fin}}} \Phi^q(T_\omega) = \infty\}.$

(2) $\inf\{q \mid \nu^{(q)}(E^\infty) = 0\} = \sup\{q \mid \nu^{(q)}(E^\infty) = \infty\}.$

(3) the unique $q > 0$ such that[d]

$$\lim_{\ell \to \infty} \left(\rho \left(\left[\sum_{\omega \in E_{ij}^\ell} \Phi^q(T_\omega) \right]_{1 \le i,j \le n} \right) \right)^{1/\ell} = 1. \tag{10}$$

We denote the common value by $\dim_{\mathrm{aff}} G_{(\Omega_1, \ldots, \Omega_n)}$ *and call it the* affinity dimension *(or Falconer dimension) of the GIFS.* □

By a covering argument, we get an upper bound for the Hausdorff dimension of the sets Ω_i, compare Proposition 5.1 of Ref. 2 and Theorem 9.12 of Ref. 3.

Proposition 4.2. *If $\nu^{(q)}(E^\infty) < \infty$, then $h^{(q)}(\Omega_i) < \infty$ for all $1 \le i \le n$. In particular, we have $\dim_{\mathrm{Hd}} \Omega_i \le \dim_{\mathrm{aff}} G_{(\Omega_1, \ldots, \Omega_n)}$ for all $1 \le i \le n$.* □

In general, it is difficult to decide whether equality holds in this last inequality for a self-affine GIFS, although in a certain sense equality is the generic case – at least in $\mathbb{R}^r$ (see Theorem 5.3 of Ref. 2 and Theorem 9.12 of Ref. 3). And contrary to the well-studied self-similar case (where $\alpha_1 = \ldots = \alpha_r$) in $\mathbb{R}^r$, even the *open set condition*[e] (OSC) does not ensure the equality sign (cf. Ref. 11 and Examples 9.10 & 9.11 in Ref. 3).

We now define a second singular value function $\Psi^q(T)$ of $T \in \mathcal{L}'$ for $q \ge 0$ as follows[4,13]:

$$\Psi^q(T) = \begin{cases} 1 & \text{if } q = 0 \\ \alpha_{r+2s+k} \cdot \ldots \cdot \alpha_{r+2s+k-j+1} \cdot \alpha_{r+2s+k-j+1}^{q-j+1} & \text{if } j - 1 < q \le j \\ (\alpha_1 \cdot \alpha_2 \cdot \ldots \cdot \alpha_{r+2s+k})^{q/(r+2s+k)} & \text{if } q > r + 2 \cdot s + k \end{cases} \tag{11}$$

Again, $\Psi^q(T) = [\Phi^q(T^{-1})]^{-1}$ is continuous and strictly decreasing in q, but super-multiplicative for fixed q. Just as in Proposition 4.1, we define the *lower affinity dimension*

$$\underline{\dim}_{\mathrm{aff}} G_{(\Omega_1, \ldots, \Omega_n)} = \inf\{q \mid \sum_{\omega \in E^{\mathrm{fin}}} \Psi^q(T_\omega) < \infty\} = \sup\{q \mid \sum_{\omega \in E^{\mathrm{fin}}} \Psi^q(T_\omega) = \infty\} \tag{12}$$

of the GIFS. Then, with the help of the "mass distribution principle" (see Proposition 4.2 in Ref. 3), we obtain the following lower bound for the Hausdorff dimension of the sets Ω_i, compare Proposition 2 of Ref. 4.

---

[d]As a reminder: $\rho(\boldsymbol{F})$ denotes the spectral radius of the matrix $\boldsymbol{F}$.
[e]The OSC is satisfied if there exist disjoint non-empty bounded open sets $(U_1, \ldots, U_n)$ such that $U_i \supset \bigcup_{i=1}^n \bigcup_{f_{ij}^{(\ell)} \in F_{ij}} f_{ij}^{(\ell)}(U_j)$, with the unions disjoint.

Proposition 4.3. *Let* $(\Omega_1, \ldots, \Omega_n)$ *be the solution of a (strongly connected) GIFS* $\Omega_i = \bigcup_{i=1}^{n} \bigcup_{f_{ij}^{(\ell)} \in F_{ij}} f_{ij}^{(\ell)}(\Omega_j)$, *where all unions are disjoint. If the sets* $(\Omega_1, \ldots, \Omega_n)$ *are also pairwise disjoint, then* $\underline{\dim}_{\mathrm{aff}} G_{(\Omega_1, \ldots, \Omega_n)} \leq \dim_{\mathrm{Hd}} \Omega_i$ *for all* $1 \leq i \leq n$. $\square$

We remark that this disjointness condition is often easy to check in the cases we are interested in, since p-adic spaces are totally disconnected.

If the linear part of all maps $f_{ij}^{(\ell)}$ is the same, i.e., $T = T_{f_{ij}^{(\ell)}}$ for all i, j, ℓ, we finally obtain the following theorem.

Theorem 4.1. *For a (strongly connected) GIFS with (unique non-empty compact) solution* $(\Omega_1, \ldots, \Omega_n)$, *where all maps* $f_{ij}^{(\ell)}$ *have the same linear part* T, *the affinity dimension* $\dim_{\mathrm{aff}} G_{(\Omega_1, \ldots, \Omega_n)}$ *is given by the unique value* $q > 0$ *such that* $\Phi^q(T) \cdot \rho(F) = 1$. *The Hausdorff dimension of the sets* Ω_i *is bounded by the affinity dimension of the GIFS, i.e.,* $\dim_{\mathrm{Hd}} \Omega_i \leq \dim_{\mathrm{aff}} G_{(\Omega_1, \ldots, \Omega_n)}$ *for all* $1 \leq i \leq n$. *Furthermore, if the unions in the GIFS are disjoint and the sets* $(\Omega_1, \ldots, \Omega_n)$ *are pairwise disjoint, the Hausdorff dimension of the sets* Ω_i *is bounded from below by the lower affinity dimension of the GIFS, i.e.,* $\underline{\dim}_{\mathrm{aff}} G_{(\Omega_1, \ldots, \Omega_n)} \leq \dim_{\mathrm{Hd}} \Omega_i$ *for all* $1 \leq i \leq n$, *where* $\underline{\dim}_{\mathrm{aff}} G_{(\Omega_1, \ldots, \Omega_n)}$ *is given by the unique value* $q > 0$ *such that* $\Psi^q(T) \cdot \rho(F) = 1$. $\square$

5. An Example

Our motivation for this work are so-called "Rauzy fractals"[15], which are used to prove pure pointedness of the dynamical system of certain 1-dimensional sequences over a finite alphabet, obtained by a substitution rule. "Rauzy fractals" yield a geometric representation[14] (or so-called windows for models sets[1]) for such sequences.

Here, we look at the substitution $a \mapsto aaba$, $b \mapsto aa$ (we obtain a two-sided infinite sequence by applying the substitution repeatedly (we denote the zeroth position by $|$): $a \,|\, a \mapsto aaba \,|\, aaba \mapsto \ldots aaaba \,|\, aabaaabaaaaaba \ldots$). From such a substitution, one can obtain a GIFS (see the above literature[15,14,1] and references therein), in this case in the space $\mathbb{R} \times \mathbb{Q}_2$:

$$\Omega_a = T(\Omega_a) \cup T(\Omega_b) \cup T(\Omega_a) + \frac{1}{2}t_1 \cup T(\Omega_b) + \frac{1}{2}t_1 \cup T(\Omega_a) + t_2$$
$$\Omega_b = T(\Omega_a) + t_1 \tag{13}$$

where $T((x_1, x_2)) = (\kappa \cdot x_1, \lambda \cdot x_2)$, $t_1 = (\kappa, \lambda)$, $t_2 = (\kappa+1, \lambda+1)$, $\kappa = \frac{3-\sqrt{17}}{2} \approx -0.562$ and $\lambda = \frac{3+\sqrt{17}}{2} \approx 3.562$, which in the 2-adic expansion starts as $\lambda = 01101\ldots$. We have $|\kappa| = \frac{2}{3}$, $\|\lambda\|_2 = \frac{1}{2}$ and $\rho(F) = \lambda$, and therefore the affinity dimension[f] $\dim_{\mathrm{aff}} G_{(\Omega_a, \Omega_b)} = 2 = \dim_{\mathrm{metr}} \mathbb{R} \times \mathbb{Q}_2$. Indeed, one can show that the Haar measure of the sets Ω_a and Ω_b is positive and the intersection $\Omega_a \cap \Omega_b$ has Haar measure 0.

---

[f]We also have $\underline{\dim}_{\mathrm{aff}} G_{(\Omega_a, \Omega_b)} = 2$, but the sets Ω_a and Ω_b are not disjoint. Therefore Proposition 4.3 does not apply here.

It is more interesting to calculate the Hausdorff dimension of the boundaries $\partial\Omega_a$ and $\partial\Omega_b$. For the boundary, one can also derive a GIFS with the same contraction T. This is possible, because the above GIFS for (Ω_a, Ω_b) can be dualised[9] to obtain a point set equation for point sets (X_a, X_b):

$$X_a = T^{-1}(X_a) \cup T^{-1}(X_a) + T^{-1}(\frac{1}{2}t_1) \cup$$
$$T^{-1}(X_b) + T^{-1}(t_1) \cup T^{-1}(X_a) + T^{-1}(t_2) \tag{14}$$
$$X_b = T^{-1}(X_a) \cup T^{-1}(X_a) + T^{-1}(\frac{1}{2}t_1)$$

where $T^{-1}((x_1, x_2)) = (\frac{1}{\kappa} \cdot x_1, \frac{1}{\lambda} \cdot x_2)$. Starting this iteration with $X_a = \{(0,0)\} = X_b$, one obtains a fixed point for (X_a, X_b) and one can show that $J = (X_a + \Omega_a) \cup (X_b + \Omega_b)$ is a tiling with the prototiles Ω_a and Ω_b of the whole space $\mathbb{R} \times \mathbb{Q}_2$ (for purely Euclidean spaces, this is now well established[8]). With the help of this tiling J, one obtains the following GIFS for the boundary:

$$
\begin{aligned}
\Xi_{(a,b,0)} &= T(\Xi_{(a,a,1)}) \\
\Xi_{(b,a,0)} &= T(\Xi_{(a,a,-1)}) + t_1 \\
\Xi_{(a,a,1)} &= T(\Xi_{(a,a,-1)}) + t_1 \cup T(\Xi_{(a,a,\frac{\lambda}{2}-1)}) \cup T(\Xi_{(b,a,\frac{\lambda}{2}-1)}) \\
\Xi_{(a,a,-1)} &= T(\Xi_{(a,a,1)}) \cup T(\Xi_{(a,a,1-\frac{\lambda}{2})}) + \tfrac{1}{2}t_1 \cup T(\Xi_{(a,b,1-\frac{\lambda}{2})}) + \tfrac{1}{2}t_1 \\
\Xi_{(a,a,\frac{\lambda}{2}-1)} &= T(\Xi_{(a,a,1)}) + \tfrac{1}{2}t_1 \\
\Xi_{(a,a,1-\frac{\lambda}{2})} &= T(\Xi_{(a,a,-1)}) + t_2 \\
\Xi_{(a,b,1-\frac{\lambda}{2})} &= T(\Xi_{(a,a,\frac{\lambda}{2}-1)}) \cup T(\Xi_{(b,a,\frac{\lambda}{2}-1)}) \\
\Xi_{(b,a,\frac{\lambda}{2}-1)} &= T(\Xi_{(a,a,1-\frac{\lambda}{2})}) + t_1 \cup T(\Xi_{(a,b,1-\frac{\lambda}{2})}) + t_1
\end{aligned}
\tag{15}
$$

Here, $\Xi_{(a,a,1-\frac{\lambda}{2})} = \Omega_a \cap \Omega_a + (1 - \frac{\kappa}{2}, 1 - \frac{\lambda}{2})$ and similarly for the other sets. The boundaries are therefore given by

$$\partial\Omega_a = \Xi_{(a,b,0)} \cup \Xi_{(a,a,1)} \cup \Xi_{(a,a,-1)} \cup \Xi_{(a,a,\frac{\lambda}{2}-1)} \cup \Xi_{(a,a,1-\frac{\lambda}{2})} \cup \Xi_{(a,b,1-\frac{\lambda}{2})}$$
$$\partial\Omega_b = \Xi_{(b,a,0)} \cup \Xi_{(b,a,\frac{\lambda}{2}-1)}. \tag{16}$$

To obtain a strongly connected GIFS which fulfills the disjointness condition from the GIFS in Eq. (15), we observe that $\Xi_{(a,b,0)} = \Xi_{(b,a,0)}$, $\Xi_{(a,a,1)} = \Xi_{(a,a,1-\frac{\lambda}{2})} \cup \Xi_{(a,b1-\frac{\lambda}{2})}$ and $\Xi_{(a,a,-1)} = \Xi_{(a,a,\frac{\lambda}{2}-1)} \cup \Xi_{(a,b,0)}$. So we arrive at the GIFS

$$
\begin{aligned}
\Xi_{(a,b,0)} &= T(\Xi_{(a,a,1-\frac{\lambda}{2})}) \cup T(\Xi_{(a,b,1-\frac{\lambda}{2})}) \\
\Xi_{(a,a,\frac{\lambda}{2}-1)} &= T(\Xi_{(a,a,1-\frac{\lambda}{2})}) + \tfrac{1}{2}t_1 \cup T(\Xi_{(a,b,1-\frac{\lambda}{2})}) + \tfrac{1}{2}t_1 \\
\Xi_{(a,a,1-\frac{\lambda}{2})} &= T(\Xi_{(a,a,\frac{\lambda}{2}-1)}) + t_2 \cup T(\Xi_{(a,b,0)}) + t_2 \\
\Xi_{(a,b,1-\frac{\lambda}{2})} &= T(\Xi_{(a,a,\frac{\lambda}{2}-1)}) \cup T(\Xi_{(b,a,\frac{\lambda}{2}-1)}) \\
\Xi_{(b,a,\frac{\lambda}{2}-1)} &= T(\Xi_{(a,a,1-\frac{\lambda}{2})}) + t_1 \cup T(\Xi_{(a,b,1-\frac{\lambda}{2})}) + t_1.
\end{aligned}
\tag{17}
$$

For this GIFS, the spectral radius $\rho(F)$ equals 2. Consequently, we obtain $\dim_{\text{aff}} G_{(\Xi_{(a,b,0)}, \Xi_{(a,a,\frac{\lambda}{2}-1)}, \ldots, \Xi_{(b,a,\frac{\lambda}{2}-1)})} = \frac{\log(\sqrt{17}-3)}{\log 2} + 1 \approx 1.1675$ and

$\underline{\dim}_{\mathrm{aff}} G_{(\Xi_{(a,b,0)},\Xi_{(a,a,\frac{\lambda}{2}-1)},\dots,\Xi_{(b,a,\frac{\lambda}{2}-1)})} = 1$. Using the total disconnectedness of $\mathbb{Q}_2$, one can show that the disjointness condition for the sets in Eq. (17) holds, wherefore these are the upper and lower bounds for the Hausdorff dimension of the boundaries $\partial\Omega_a$ and $\partial\Omega_b$. We end this article with pictures of the GIFS in Eq. (17) and of the sets Ω_a, Ω_b and their boundaries.
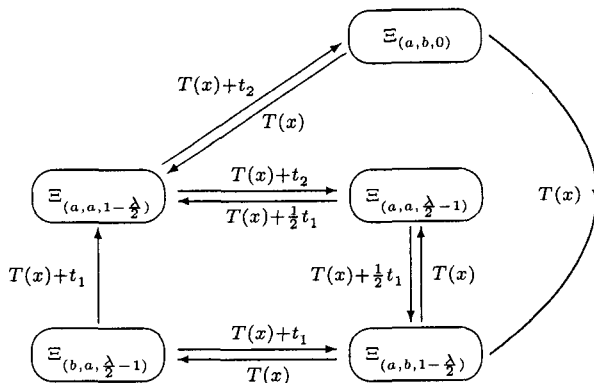

Figure 1. The directed graph $G_{(\Xi_{(a,b,0)},\Xi_{(a,a,\frac{\lambda}{2}-1)},\dots,\Xi_{(b,a,\frac{\lambda}{2}-1)})}$ associated to the GIFS in Eq. (17).

Acknowledgments

References

1. M. Baake and B. Sing, "Kolakoski-(3,1) is a (deformed) model set", *Canad. Math. Bull.* **47** (2004) 168–190; math.MG/0206098.
2. K. Falconer, "The Hausdorff dimension of self-affine fractals", *Math. Proc. Camb. Phil. Soc.* **103** (1988) 339–350.
3. K. Falconer, *Fractal Geometry*, John Wiley & Sons, Chicester, 1990.
4. K. Falconer, "The Hausdorff dimension of self-affine fractals II", *Math. Proc. Camb. Phil. Soc.* **111** (1992) 169–179.
5. P.R. Halmos, *Measure Theory*, Springer, New York, 1974.
6. E. Hewitt and K.A. Ross, *Abstract Harmonic Analysis I*, Springer, Berlin, 1963.
7. W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton University Press, Princeton, NJ, 1948.
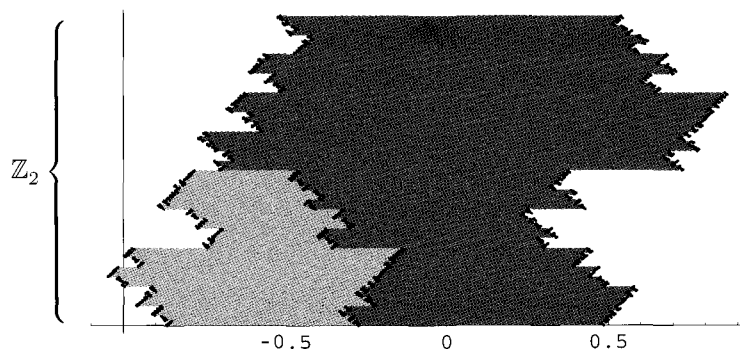
Figure 2. The sets Ω_a (dark gray) and Ω_b (light gray) and their boundaries (black) in $\mathbb{R} \times \mathbb{Q}_2$.

8. S. Ito and H. Rao, "Atomic surfaces, tilings and coincidence I. Irreducible case", Preprint.

9. J.C. Lagarias and Y. Wang, "Substitution Delone sets", *Discr. Comput. Geom.* **29** (2003) 175–209.

10. S. Lang, *Algebraic Number Theory*, Addison-Wesley, Reading, MA, 1970.

11. C. McMullen, "The Hausdorff dimension of general Sierpiński carpets", *Nagoya Math. J.* **96** (1984) 1–9.

12. M.E. Munroe, *Measure and Integration*, Addison-Wesley, Reading, MA, 1971.

13. W.H. Paulsen, "Lower bounds for the Hausdorff dimension of n-dimensional self-affine sets", *Chaos, Solitons & Fractals* **5** (1995) 909–931.

14. N. Pytheas-Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, (*Lect. Notes Math.* **1784**, edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel), Springer, Berlin, 2002.

15. G. Rauzy, "Nombres algébriques et substitutions", *Bull. Soc. math. France* **110** (1982) 147–178.

16. A.M. Robert, *A Course in p-adic Analysis*, Springer, New York, 2000.

17. C.A. Rogers, *Hausdorff Measures*, Cambridge University Press, Cambridge, 1998 (Reissue of the 1970 edition).

18. J.-P. Serre, *Local Fields*, Springer, New York, 1979.

MULTIFRACTALITY OF THE MULTIPLICATIVE AUTOREGRESSIVE POINT PROCESSES

B. KAULAKYS, M. ALABURDA, V. GONTIS AND T. MESKAUSKAS

Institute of Theoretical Physics and Astronomy of Vilnius University,
A. Gostauto 12,
LT-01108 Vilnius, Lithuania
E-mail: kaulakys@itpa.lt

Multiplicative processes and multifractals have earned increased popularity in applications ranging from hydrodynamic turbulence to computer network traffic, from image processing to economics. We analyse the multifractality of the recently proposed point process models generating the signals exhibiting $1/f^\beta$ noise. The models may be used for modeling and analysis of stochastic processes in different systems. We show that the multiplicative point process models generate multifractal signals, in contrast to the formally constructed signals with $1/f^\beta$ noise and signals consisting of sum of the uncorrelated components with a wide-range distribution of the relaxation times.

1. Introduction

Multifractal models are used to account for scale invariance properties of various objects in different domains ranging from the energy dissipation in turbulent flows[1] to financial data[2]. Healthy human heartbeat intervals exhibit multifractal properties rather than being fractal for a life-threatening condition, known as congestive heart failure[3]. Cerebral blood flow in healthy humans is also multifractal[4].

Scaling behavior has become a welcome careful description of complexity in many fields including natural phenomena, human heart rhythm in biology, spatial repartition of faults in geology, as well as human activities such as traffic in computer networks and financial markets. The multifractal formalism has received much attention as one of the most popular frameworks to describe and analyse signals and processes that exhibit scaling properties, covering and connecting both the local scaling and the global one in terms of sample moments.

The purpose of this paper is to analyse the multifractality of signals exhibiting $1/f^\beta$ noise generated by different techniques and, especially, of the point processes with $1/f^\beta$ power spectral density[5,6].

First of all, however, we will analyse the multifractality of the signal constructed by the inverse fast Fourier transform[7]. Using this method we can generate signals with any desirable slope β of the power spectral density $S(f) \sim 1/f^\beta$.

We calculate a generalized qth order height-height correlation function (GHCF)

$F_q(t)$ defined as[8]

$$F_q(t) = \langle |I(t' + t) - I(t')|^q \rangle^{1/q},\qquad(1)$$

where the angular brackets denote the time average. The GHCF $F_q(t)$ characterizes the correlation properties of the signal $I(t)$, and for a multiaffine signal a power-law behavior like

$$F_q(t) \sim t^{H_q}\qquad(2)$$

is expected. Here H_q is the generalized qth order Hurst exponent. If H_q is independent on q, a single scaling exponent H_q is involved and the signal $I(t)$ is said to be monofractal[8]. If H_q depends on q, the signal is considered to be multifractal.
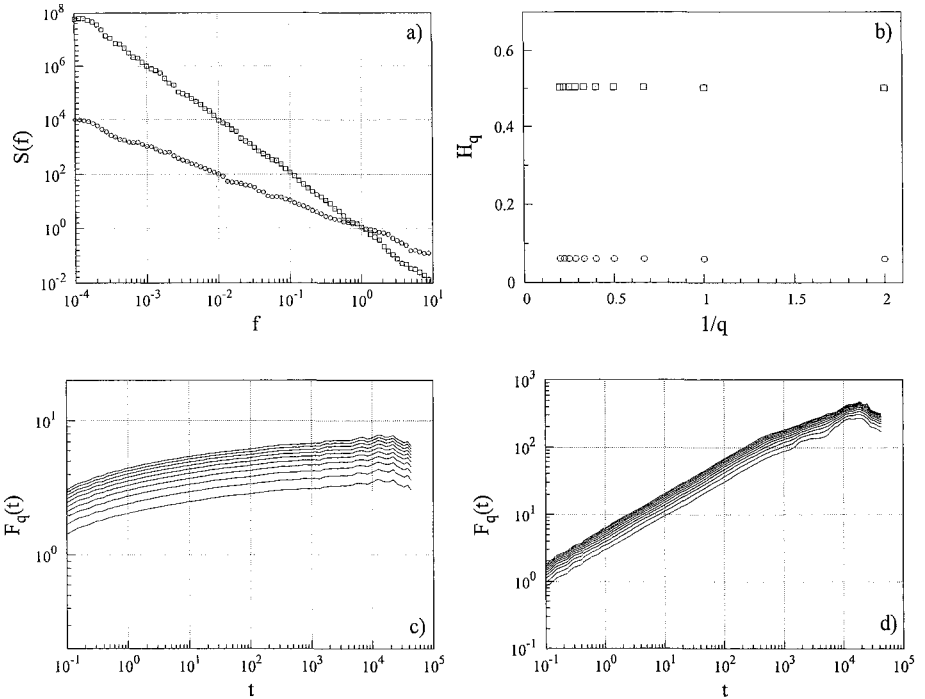
Figure 1. a) Power spectral density and b) the generalized Hurst exponents H_q versus $1/q$ in the scaling regime $1 < t < 1000$ for the slopes $\beta = 1$ (open circles) and $\beta = 2$ (open squires). The signal of 10^6 points was generated and averaged over 10 realizations. c) and d) show GHCF $F_q(t)$ versus time t for the same parameters $\beta = 1$ and $\beta = 2$, respectively.

In figure 1 a) we present a power spectral densities with the different slopes β and in figure 1 b) we show the Hurst exponents, calculated from GHCF using linear regression dependence on $1/q$ of the signals formally constructed by the inverse Fourier transform. In figure 1 c) and d) corresponding GHCF $F_q(t)$ versus time t are shown. We see that Hurst exponent H_q does not depend on q, which indicates that the signal is monofractal.

2. Stochastic multiplicative point process

In many cases the intensity of some signals or currents can be represented by a sequence of random (however, as a rule, mutually correlated) pulses or elementary events $A_k(t - t_k)$,

$$I(t) = \sum_k A_k(t - t_k). \tag{3}$$

Here the function $A_k(\phi)$ represents the shape of the k pulse making an influence on the signal $I(t)$ in the region of the transit time t_k. We will be interested in the processes with the power-law distribution of the power spectral density at low frequencies. It is easy to show that the shapes of the pulses mainly influence the high frequency, $f \gtrsim 1/\Delta t_p$, with Δt_p being the characteristic pulse length, power spectral density, while fluctuations of the pulse amplitudes result, as a rule, in the white or the Lorentzian but not $1/f$ noise[9]. Therefore, we restrict our analysis to noise due to correlations between the transit times t_k. In such approach we can replace the function $A_k(t - t_k)$ by the Dirac delta function and then express the signal as

$$I(t) = \bar{a} \sum_k \delta(t - t_k), \tag{4}$$

with $\bar{a}$ being an average contribution to the signal of one pulse. This model[5] also corresponds to the flow of identical objects: electrons, photons, cars, and so on, and is called the point process model. Point processes arise in different fields, such as physics, economics, cosmology, ecology, neurology, seismology, traffic flow, signaling and telecom networks, and the Internet (see e.g., papers[9,10] and references herein).

The power spectrum of the point process signal is described completely by the set of the interevent intervals $\tau_k = t_{k+1} - t_k$. Moreover, the low frequency noise is defined by the statistical properties of the signal at a large-time-scale, i.e., by the fluctuations of the time difference

$$\Delta(k; q) \equiv t_{k+q} - t_k = \sum_{i=k}^{k+q-1} \tau_i \tag{5}$$

at large q, determined by the slow dynamics of the average interpulse time $\tilde{\tau}_k(q) = \Delta(k; q)/q$ between the occurrence of pulses k and $k + q$. Quite generally the dependence of the average interevent time $\tilde{\tau}_k$ may be described by the general Langevin equation. The Langevin equation may be written down in the actual time t or, equivalently, in the space of the occurrence numbers k with the drift coefficient $h(\tilde{\tau}_k)$ and a multiplicative noise $g(\tilde{\tau}_k)\xi(k)$,

$$\frac{d\tilde{\tau}_k}{dk} = h(\tilde{\tau}_k) + g(\tilde{\tau}_k)\xi(k). \tag{6}$$

Here we interpret k as a continuous variable while the white Gaussian noise $\xi(k)$ satisfies the standard condition

$$\langle \xi(k)\xi(k') \rangle = \delta(k - k') \tag{7}$$

with the brackets $\langle \ldots \rangle$ denoting the averaging over the realizations of the process. We understand the equation (6) in Itô interpretation.

Transition from the occurrence numbers k to the actual time t in Eq. (6) may be fulfilled using the relation $dt = \tilde{\tau}_k dk$[11].

The particular sequence of the interevent times τ_k may be superimposed by some additional noise or stochasticity, e.g., τ_k may be determined by the Poisson distribution

$$P(\tau_k) = \frac{1}{\tilde{\tau}_k} e^{-\tau_k / \tilde{\tau}_k} \tag{8}$$

with the slowly, according to Eq. (6), changeable average interevent time $\tilde{\tau}_k$. Such additional stochasticity do not influence the long-range statistical properties and the low frequency spectra of the process. Therefore, further we will restrict the analysis to the processes generated by Eq. (6) and will identify τ_k with $\tilde{\tau}_k$.

2.1. Power spectral density

The point process is entirely defined by the occurrence times t_k. The power spectral density of the point process (4) may be expressed as

$$S(f) = \lim_{T \to \infty} \left\langle \frac{2}{T} \left| \int_{t_i}^{t_f} I(t) e^{-i2\pi f t} dt \right|^2 \right\rangle = \lim_{T \to \infty} \left\langle \frac{2\bar{a}^2}{T} \sum_k \sum_{q=k_{\min}-k}^{k_{\max}-k} e^{i2\pi f \Delta(k;q)} \right\rangle, \tag{9}$$

where t_i and t_f are initial and final observation times, $T = t_f - t_i \gg \omega^{-1}$ is the whole observation time and $\omega = 2\pi f$. Here $k_{\min}$ and $k_{\max}$ are minimal and maximal values of index k in the interval of observation T and the brackets $\langle \ldots \rangle$ denote the averaging over realizations of the process.

For the interpulse intervals described by the Langevin equation (6) we use a perturbative solution in the vicinity of τ_k. After replacing the averaging over k by the averaging over the distribution $P_k(\tau_k)$ of the interpulse times τ_k, we have the power spectrum[6]

$$S(f) = 2\bar{I}^2 \frac{\bar{\tau}}{\sqrt{\pi f}} \int_0^\infty P_k(\tau_k) \operatorname{Re}[e^{-i(x - \frac{\pi}{4})} \operatorname{erfc} \sqrt{-ix}] \frac{\sqrt{x}}{\tau_k} d\tau_k, \tag{10}$$

where $\bar{I}$ and $\bar{\tau}$ are the averages of the signal and the interpulse times, respectively, and $x = \pi f \tau_k^2 / h(\tau_k)$.

The replacement of the averaging over k and over realizations of the process by the averaging over the distribution of the interpulse times τ_k, $P_k(\tau_k)$, is possible when the process is ergodic. Ergodicity is usually a common feature of the stationary process described by the general Langevin equation.

According to Eq. (10) the small interpulse times and the clustering of the pulses make the greatest contribution to $1/f^\beta$ noise. The power-law spectral density is very often related with the power-law behavior of other characteristics of the signal,

such as autocorrelation function, probability densities and other statistics, and with the fractality of the signals, in general[12]. Therefore, we investigate the power-law dependences of the drift coefficient and of the distribution density on the time τ_k in some interval of the small interpulse times, i.e.,

$$h(\tau_k) = \gamma\tau_k^\delta, \quad P_k(\tau_k) = C\tau_k^\alpha, \quad \tau_{\min} \leq \tau_k \leq \tau_{\max}, \tag{11}$$

where the coefficient γ represents the rate of the signal's nonlinear relaxation and C has to be defined from the normalization.

The simplest and the well-known process generating the power-law probability distribution function for τ_k is a multiplicative stochastic process with $g(\tau_k) = \sigma\tau_k^\mu$ and $\delta = 2\mu - 1$, written as[6,13]

$$\tau_{k+1} = \tau_k + \gamma\tau_k^{2\mu-1} + \sigma\tau_k^\mu\varepsilon_k. \tag{12}$$

Here γ represents the nonlinear relaxation of the signal, while τ_k fluctuates due to the perturbation by normally distributed uncorrelated random variables ε_k with a zero expectation and unit variance and σ is a standard deviation of the white noise.

Eq. (12) is the difference (discrete) version of the differential equation (6). On the other hand, it is the generalization of the simple autoregressive model of $1/f$ noise[5] (see also Eq. (15)) and represents quite general evolution of the interevent time with the nonlinear drift $h(\tau_k) = \gamma\tau_k^{2\mu-1}$ and the multiplicative noise $\sigma\tau_k^\mu\varepsilon_k$, resulting in the $1/f^\beta$ noise and power-law distribution (11) of the interevent time τ_k with the exponent $\alpha = 2\gamma/\sigma^2 - 2\mu$. Indeed, the power spectrum for the process (12), when $\gamma/(\pi\tau_{\max}^{2-\delta}) \ll f \ll \gamma/(\pi\tau_{\min}^{2-\delta})$, is[6]

$$S(f) = \frac{(2+\alpha)(\beta-1)\bar{a}^2\Gamma(\beta-1/2)}{\sqrt{\pi}\alpha(\tau_{\max}^{2+\alpha} - \tau_{\min}^{2+\alpha})\sin(\pi\beta/2)}\left(\frac{\gamma}{\pi}\right)^{\beta-1}\frac{1}{f^\beta}, \tag{13}$$

where

$$\alpha = \frac{2\gamma}{\sigma^2} - 2\mu, \quad \beta = 1 + \frac{\alpha}{3-2\mu}, \quad \frac{1}{2} < \beta < 2. \tag{14}$$

For $\mu = 1$ we have a completely multiplicative point process when the stochastic change of the interpulse time is proportional to itself. Another case of interest concerns $\mu = 1/2$, then we have the Brownian motion of the interevent time with the linear relaxation of the signal $I \simeq \bar{a}/\tau$.

Figure 2 represents the spectral densities (9) with different slopes β of the signals generated numerically according to Eqs. (4) and (12) for different parameters of the model. We see that the simple iterative equation (12) with the multiplicative noise produces the signals with the power spectral density of different slopes, depending on the parameters of the model. The agreement of the numerical results with the approximate theory is quite good.
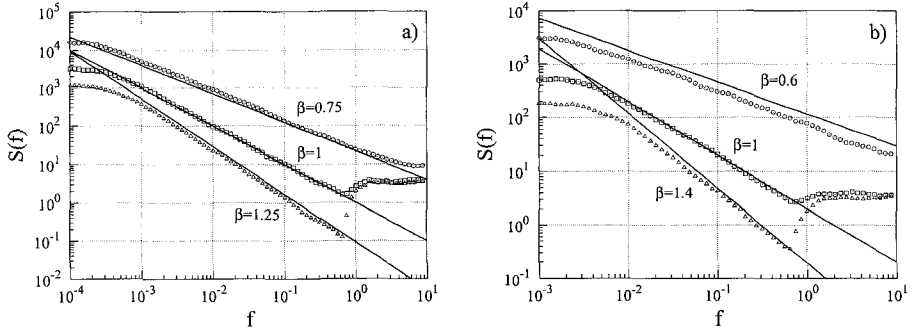
Figure 2. Power spectral density (9) vs frequency of the signal generated by Eqs. (4) and (12) with the parameters a) $\mu = 0.5$, $\sigma = 0.02$ and different relaxations of the signal $\gamma = 0.0001$ (open circles), 0.0002 (open squares) and 0.0003 (open triangles); and b) $\mu = 1$, $\sigma = 0.1$ and $\gamma = 0.008$ (open circles), 0.01 (open squares) and 0.012 (open triangles). We restrict the diffusion of the interevent time in the interval $\tau_{\min} = 10^{-6}$, $\tau_{\max} = 1$ with the reflective boundary condition at $\tau_{\min}$ and transition to the white noise, $\tau_{k+1} = \tau_{\max} + \sigma \varepsilon_k$, for $\tau_k > \tau_{\max}$ and 100 realizations with 10^6 t_k points each were used. The solid lines represent the analytical results according to Eq. (13).

2.2. Multifractal point processes

The multifractal formalism has received much attention recently as one of the most popular frameworks to describe and analyse signals and processes that exhibit the scaling properties.

Fractality of the point process can be investigated by transition from the point process to the stochastic signal $I(t)$, using the rectangular constant area pulses, instead of the Dirac delta functions. The stochastic signal will have the same fractal properties as the origin point process.
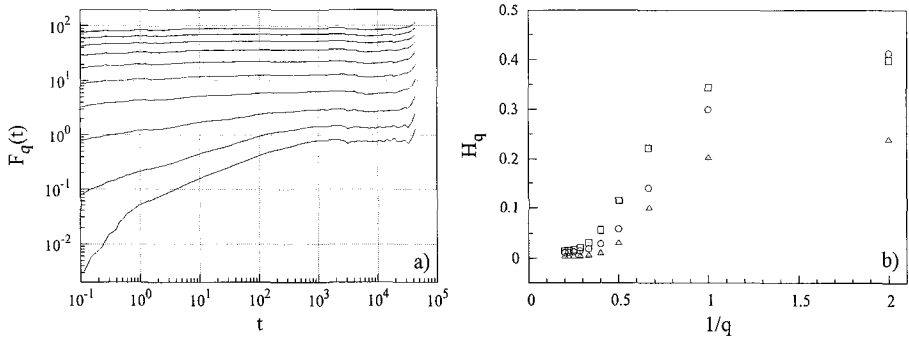
Figure 3. a) Generalized height-height correlation function $F_q(t)$ versus time t for the signal (4) and (12) with $\mu = 0.5$, $\bar{\tau} = 1$, $\sigma = 0.02$, $\gamma = 0.0002$, and $q = 0.5, 1, 1.5, \ldots, 5$ from bottom to top. b) The generalized Hurst exponents H_q versus $1/q$ in the scaling regime $1 < t < 1000$ for the $\bar{\tau} = 1$ and $\mu = 0.5$, $\sigma = 0.02$, $\gamma = 0.0002$ (open circles); $\mu = 0.5$, $\sigma = 0.02$, $\gamma = 0.0003$ (open squares); $\mu = 1$, $\sigma = 0.1$, $\gamma = 0.008$ (open triangles).

In figure 3 a) we present the GHCF as a function of the time interval t, and in figure 3 b) we show the Hurst exponents calculated from GHCF using the linear regression dependence on $1/q$ for different parameters μ, σ, and γ. We observe the clear multifractal behavior since the slopes of the log-log plot of GHCF are depending on q.

Another interesting case is a sequence of transit times with random increments of the time intervals between pulses, $\tau_k = \tau_{k-1} + \sigma\varepsilon_k$. It is natural to restrict in some way the infinite Brownian increase or decrease of the interpulse times τ_k, e.g., by the introduction of the relaxation to the average interpulse time $\bar{\tau}$ rate γ. So, we have an additive point process

$$\tau_k = \tau_{k-1} - \gamma(\tau_{k-1} - \bar{\tau}) + \sigma\varepsilon_k. \tag{15}$$

This model generates the process with $1/f$ noise[5], and may be useful for modeling and analysing different systems (see references in paper[6]). Introduction of the reflective boundary condition at $\tau_{min} > 0$ avoids the formation of clusters and leads to $1/f^2$ noise.
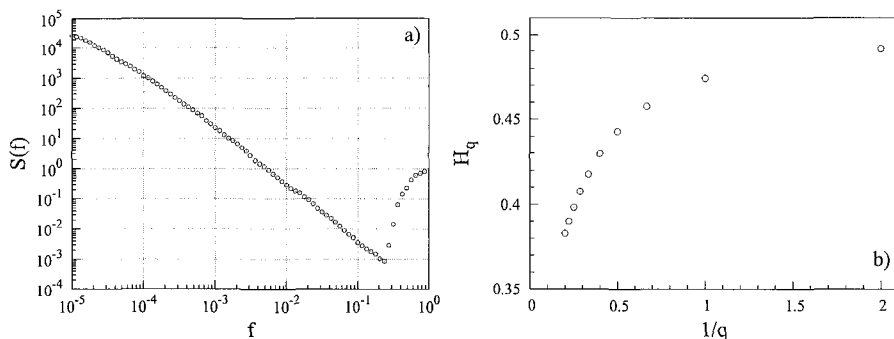
Figure 4.   a) Power spectral density and b) the generalized Hurst exponents of the additive point process (15) with the reflecting boundary condition at $\tau_{min} = 0.1$ in the scaling regime $1 < t < 1000$. The signal of 10^6 points was generated with the parameters $\bar{\tau} = 1$, $\sigma = 0.001$, and $\gamma = 0.000001$.

In figure 4 a) we present a power spectral density of the additive point process (15) with the reflecting boundary condition and in figure 4 b) we show the generalized Hurst exponents. We observe the multifractal behavior of the additive point process.

2.3. *Monofractality of the white and the Gaussian noises*

It is well-known[9] that the point processes with the interevent time τ_k distributed according to Poisson distribution

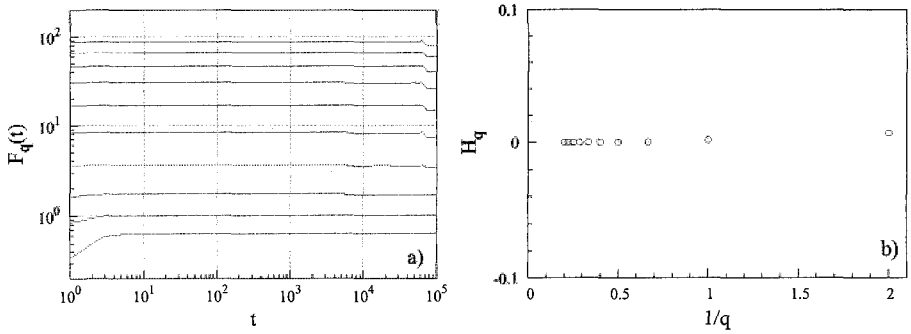$$P(\tau_k) = \frac{1}{\bar{\tau}}e^{-\tau_k/\bar{\tau}} \tag{16}$$

Figure 5. a) Generalized height-height correlation function $F_q(t)$ versus time t, and b) the generalized Hurst exponents H_q versus $1/q$ of the point process with interevent time τ_k distributed according to Poisson distribution (16) in the scaling regime $1 < t < 1000$. The signal of 10^6 points with the parameter $\bar{\tau} = 1$ has been generated.

generate the white noise, $S(f) = const$.

In figure 5 a) we present the GHCF as a function of the time interval t and in figure 5 b) we show the Hurst exponents for the white noise. The Hurst exponents of the white noise are equal to zero. This demonstrates the absence of the scaling and that there is no correlation in time.

Noise with the power-law spectral density $1/f$ is often modeled as the sum of the Lorentzian spectra with the appropriate weights of a wide range distribution of the relaxation times τ^{rel}. The signal may be expressed as a sum of N uncorrelated components,

$$I(t) = \sum_{l=1}^{N} I_l(t) = \int_{\gamma_{min}}^{\gamma_{max}} I(t,\gamma)g(\gamma)d\gamma, \tag{17}$$

where $g(\gamma)$ is the distribution of the relaxation rates $\gamma = 1/\tau^{rel}$, and every component I_l satisfies the stochastic differential equation

$$\dot{I_l} = -\gamma_l(I_l - \bar{I_l}) + \sigma_l \xi_l(t). \tag{18}$$

Here $\bar{I_l}$ is the average value of the signal component I_l, $\xi_l(t)$ is the δ-correlated white noise, $\langle \xi_l(t)\xi_{l'}(t') \rangle = \delta_{l,l'}\delta(t - t')$, and σ_l is the intensity (standard deviation) of the white noise. The steady-state solution of the stationary Fokker-Planck equation corresponding to Eq. (18) for each component I_l and the resulting signal (17) yields the Gaussian distribution densities, however, the power spectrum may be of the power-law form when $\sigma^2(\gamma)g(\gamma)$ is constant or a power-law function[6].

In figure 6 a) we present a power spectral density of the sum of the signals with a wide range distribution of the relaxation times τ^{rel} and in figure 6 b) we show the Hurst exponents, calculated from GHCF using linear regression dependence on $1/q$. In the figures we observe $1/f$ behavior of the signal noise and clearly see that Hurst exponent H_q does not depend on q, which shows that the signal (17) is monofractal.
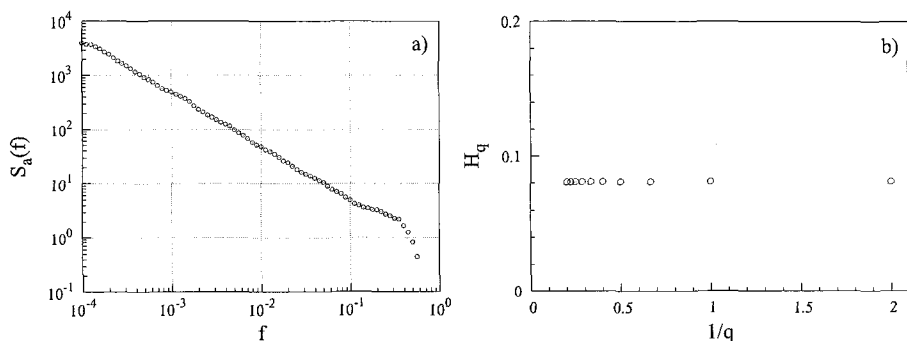
Figure 6. a) Power spectral density and b) the generalized Hurst exponents H_q versus $1/q$ in the scaling regime $1 < t < 1000$. The signal of 10^6 points was generated from 10 components with the parameters $\bar{I} = 20$, $\sigma^2(\gamma)g(\gamma) = 10$, and uniform distribution of $\lg \gamma$ with γ values in the interval $10^{-4} - 1$.

3. Conclusions

The multiplicative (12) and additive (15) stochastic point processes may generate time series exhibiting the power spectral density $S(f) \sim 1/f^\beta$ and show clear multi-fractal behavior, however, the formally constructed by the inverse Fourier transform signal, Poisson white noise (16) and Gaussian, Eqs. (17) and (18), signal with the power spectral density $S(f) \sim 1/f$ are monofractal.

Therefore, the proposed[5] and generalized[6] point process models of $1/f^\beta$ noise may be used for modeling of stochastic multifractal processes in different systems.

Acknowledgments

References

1. U. Frisch, *Turbulence: The Legacy of A. N. Kolmogorov*, Cambridge University Press (1995).
2. J.-P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivative Pricing*, Cambridge University Press (1999).
3. P. C. Ivanov, M. G. Rosenblum, L. A. N. Amaral, Z. R. Struzik, S. Havlin, A. L. Goldberger, and H. E. Stanley, Nature **399**, 461 (1999); P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger et al., Chaos **11**, 641 (2001).
4. B. J. West, M. Latka, M. Glaubic-Latka, and D. Latka, Physica A **318**, 453 (2003).
5. B. Kaulakys and T. Meškauskas, Phys. Rev. E **58**, 7013 (1998); B. Kaulakys, Phys. Lett. A **257**, 37 (1999).
6. B. Kaulakys, V. Gontis, and M. Alaburda, Phys. Rev. E **71**, 051105 (2005).
7. J. Timmer and M. Konig, Astron. Astrophys. **300**, 707 (1995).
8. P. Meakin, *Fractals, Scaling and Growth Far From Equilibrium*, Cambridge University Press (1998); E. Bacry, J. Delour, and J. F. Muzy, Phys. Rev. E **64**, 026103 (2001);

D.R. Bickel, Fractals **11**, 245 (2003); J. W. Lee, K. E. Lee, and P. A. Rikvold, arXiv: nlin/0412038 (2004).

9. T. Lukes, Proc. Phys. Soc. **78**, 153 (1961); C. Heiden, Phys. Rev. **188**, 319 (1969); K. L. Schick and A. A. Verveen, Nature **251**, 599 (1974).

10. S. Thurner, S. B. Lowen, M. C. Feurstain et al., Fractals **5**, 565 (1997); L. Telesca, V. Cuomo, V. Lapenna, and M. Macchiato, Fluct. Noise Lett. **2**, L357 (2002); T. Musha and H. Higuchi, Jap. J. Appl. Phys. **15**, 1271 (1976); A. J. Field, U. Harder, and P. G. Harrison, IEE Proceedings-Communications **151**, 355 (2004); I Csabai, J. Phys. A **27**, L417 (1994); F. Grüneis, Physica A **123**, 149 (1984); T. Musha and K. Shimizu, Jap. J. Appl. Phys. **26**, 2022 (1987); M. Y. Choi and H. Y. Lee, Phys. Rev. E **52**, 5979 (1995); F. Grüneis, Fluct. Noise Lett. 1, R119 (2001); J. M. Halley and P. Inchausti, Fluct. Noise Lett. 4, R1 (2004).

11. B. Kaulakys and J. Ruseckas, Phys Rev. E **70**, 020101 (R) (2004).

12. B. B. Mandelbrot, *Multifractals and 1/f Noise*, Springer-Verlag (1999); B. B. Mandelbrot, *Fractals and Scaling In Finance*, Springer-Verlag (1997); R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press (1999); X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley, Nature **423**, 267 (2003).

13. V. Gontis and B. Kaulakys, Physica A **343**, 505 (2004); V. Gontis and B. Kaulakys, Physica A **344**, 128 (2004).

ANALYSIS OF GEOGRAPHICAL DISTRIBUTION PATTERNS IN PLANTS USING FRACTALS

A. BARI, G. AYAD, S. PADULOSI AND T. HODGKIN

International Plant Genetic Resources Institute,

Via dei Tre Denari 472/a, Rome, Italy

A. MARTIN AND J. L. GONZALEZ-ANDUJAR

Instituto de Agricultura Sostenible (IAS-CSIC),

Apdo 4084, 14080 Córdoba, Spain

A. H. D. BROWN

Centre for Plant Biodiversity Research, CSIRO,

PO Box 1600, Canberra ACT 260, Australia

Geographical distribution patterns in plants have been observed since primeval times and have been used by plant explorers to trace the origin of plants species. These patterns embody the effects of fundamental law-like processes. Diversity in plants has also been found to be proportionate with the area, and this scaling behavior is also known as fractal behavior. In the present study, we use fractal geometry to analyze the distribution patterns of wild taxa of cowpea with the objective to locate where their diversity would be the highest to aid in the planning of targeted explorations and conservation measures.

1   Introduction

Geographical distribution patterns in plants have been recorded by explorers since primeval time and such observations have continued throughout history. The oldest known plant exploration dates back to 1495 BC by the Pharaohs in the land of Punt (Somalia today) under the reign of Queen Hatshepshut who sent the explorer Nehsi in search of plant species native to Punt Land, such as frankincense to produce aromatic gum resin and other shrubs that produce myrrh [1]. Our present knowledge of geographical patterns of variation was established in the 1920s, following N.I. Valvilov's work on centers of origin and diversity of cultivated plants [2]. These distribution patterns have been used extensively by the explorers to trace the origin and evolution of cultivated plants in the last 3 to 4 decades. Plant species exhibit these geographic distribution patterns within boundaries set up by ecological and evolutionary processes [2-5].

Common to these distribution patterns is the presence of fundamental law-like processes, which include among many others, laws of biological inheritance and evolution by natural selection [6, 7]. The relationship between the diversity and the area it occupies has been also found to follow a power-function relationship, in which the log values of diversity in terms of species' richness and area produce straight lines in a plot.

Olof Arrhenius first proposed the mathematical description of this relationship in 1920 [8] and MacArthur and Wilson later recognized the species-area relationship in 1967 [9], in their theory of island biogeography. The theory also stipulates that the species richness is a function of both the size of the area and the proximity to sources areas [9]. This relationship has been recognized in recent years in the design of *in situ* conservation or reserve areas [10, 11]

In the present project, we used fractals to study the geographical patterns of plant species distributions modeled as fractal geometrical "objects" [4] and we then measured their attributes such as size and shape. Prior to defining the different attributes for the different taxa, the geographical patterns were generated by predicting the probability distributions of where the taxon could be found based on available location records and climate data.

2 Material and Methods

2.1. *Plant species used in the study*

The present study was conducted by modeling the pattern of distribution of the wild cowpea (*Vigna unguiculata* (L.) Walpers), which encompasses the 13 wild taxa in Table 1 [12]. Some of these wild varieties are widely distributed, such as *Vigna unguiculata* subs. *dekindtiana* var. *dekindtiana*, while others are rare and confined to limited areas. Most of the varieties are located in the south-eastern part of the African continent. This area harbors the highest degree of variation among the taxa studied, and can be considered a center of diversity of cowpea wild forms [11, 12]. Interestingly, this area also falls within a biodiversity hotspot as designated by Biodiversity International [13].

Table 1: Taxa of wild cowpea and their status in terms of abundance.

Taxon	Sub-species	Variety	Status
1	*dekindtiana*	*dekindtiana*	very common
2		*ciliolate*	very rare
3		*affinis*	fairly common
4		*congolensis*	restricted
5		*grandiflora*	very rare
6		*hullensis*	restricted
7	*pubescens*		fairly common
8	*protracta*		fairly common
9		*kgalagadiensis*	fairly common
10		*rhomboidea*	restricted
11	*tenuis*		common
12		*oblonga*	common
13		*parviflora*	very rare

2.2. *Defining the extent of geographical distributions (geographical range)*

The data on wild cowpea taxa of Padulosi [12] consist of records from botanical explorations (herbarium specimens and inventories) and plant genetic resources

explorations. Each record has a number of attributes, including the geographical location (i) coordinates (x_i, y_i) where the taxon was originally sampled. From these records, the dot distribution maps were produced using the geographic coordinates (x_i, y_i) (Figure 1a). Because the dot distribution maps represent only the locations that have been previously sampled, they are likely to be biased by the locations' accessibility to explorers. Therefore we generated the probability surfaces (Figure 1b) where the taxon would likely be found based on climate data using FloraMap, a geographical information system (GIS) application developed by the International Center for Tropical Agriculture. FloraMap defines areas that are likely to contain specific desired adaptations, taxa, or habitats of interest, and areas that are potentially under-represented in the original data [14, 15].

The climate at each location (i) was extracted from a built-in climate database of FloraMap, and used as a calibration set to calculate a new climate probability model. Prior to the generation of the predicted distribution maps, calibration can be carried out to normalize rainfall pattern and determine the number (k) of orthogonal components (climate variables) through Principal Components Analysis (PCA). FloraMap then uses the Student's t-Test to calculate the probability that a map pixel belongs to the calibration based on the equation for multiple dimensions with the k independent (orthogonal) variates from the sampled data [14]:

$$z = \frac{1}{\sqrt{2\pi}} e^{-\left(t_1^2 + t_2^2 + \ldots t_k^2\right)/2} \tag{1}$$

The probability surface or grid is then generated through interpolation based on the inverse distance weighting method as per the equation:

$$z_0 = \frac{\sum_{i=1}^{n} z_i / d_i^2}{\sum_{i=1}^{n} 1 / d_i^2} \tag{2}$$

where z_0 is the interpolated value, n refers to the number of sampled plus predicted neighboring data points with its z_i value and d_i represents the distance between z_0 and z_i.

These probability surfaces are also shape-files, with each pixel defined as a square polygon that can also be used in a GIS environment to determine if there are any correlations with other environmental factors. These data can be overlaid with other information or grid data such as soils or and natural vegetation to assess the distribution of the species and monitor its distribution in space and eventually its fluctuation in time.

2.3. Geographical pattern analyses using fractals

Once the probability map for each taxon was obtained, the fractal dimensions were measured using the box counting method (D_b), the information method (D_e), and the fragmentation method (D_f) [16].

The box counting method was based on the equation:

$$N(d) \approx \frac{1}{d^{D_b}}$$ (3)

where $N(d)$ is the number of boxes and d is the box-side length. D_b is then estimated by plotting the number of boxes $N(d)$ against the box-side length d. This method is similar to the areography technique in which boxes are fixed 'sub-regions" defined by geographical coordinates [17, 18] instead of grids of changing sizes.

The information dimension (D_e) method is similar to the box counting method except that more weight is given to boxes with more points. As boxes containing a large number of points count more than boxes with fewer points, the equation is:

$$H(d) \approx -D_e \log(d)$$ (4)

where $H(d)$ is the entropy, which commonly used in information theory. This value includes both the number of boxes and their evenness or occurrence based on the number of points they contain as per the equation:

$$H = \sum p_i \log(p_i)$$ (5)

where p_i is estimated as n_i/N, n_i is the number of occurrence or evenness of the box i, and N is the total number of points across all the boxes.

In the case of fragmentation method, the fractal dimension (D_f) is defined as the exponent D_f where $N(a)$ is the number of areas or patches of size a. The distribution of the areas according to a power-law is:

$$N(a) = a^{D_f}$$ (6)

The different measurements of the fractal dimensions of all the above equations were carried out using Benoit Fractal Analysis Software version 1.3 (TruSoft International, Florida, USA) [16].

2.4. Size and shape of the taxons' patterns of distribution

To define the size and the shape of the area, we consider that the centroid of the probability surface and the variances contain information on the extent (size) of the distribution, and that the correlation between latitude and longitude provides information on the shape and orientation [4]. Both the shape and the size parameters are expressed in terms of second-order central moments as:

$$\varepsilon = \frac{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}{(\mu_{00})}$$ (7)

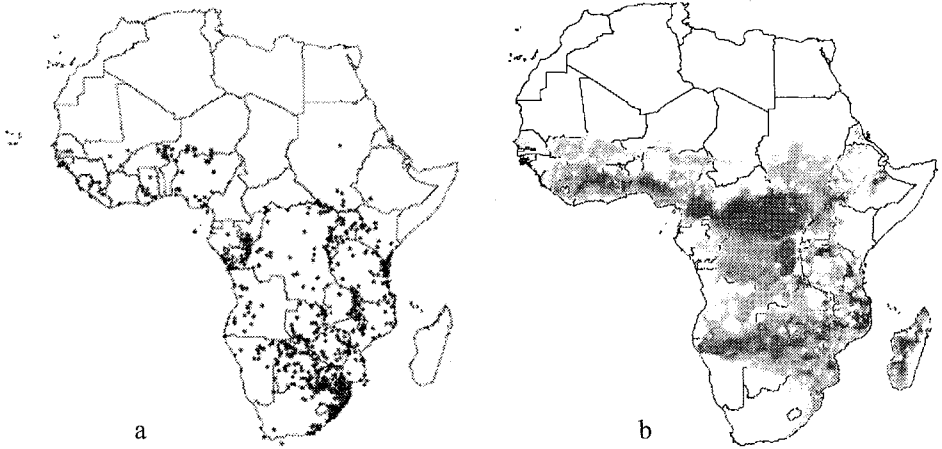$$\theta = \frac{1}{2} \arctan\left[\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right]$$ (8)

Figure 1. a) Dot distribution of 1376 locations sampled through various explorations including genetic resources records. b) Probability surface ("geographical range") for the wild cowpea varieties based on dot distribution data and climate data.
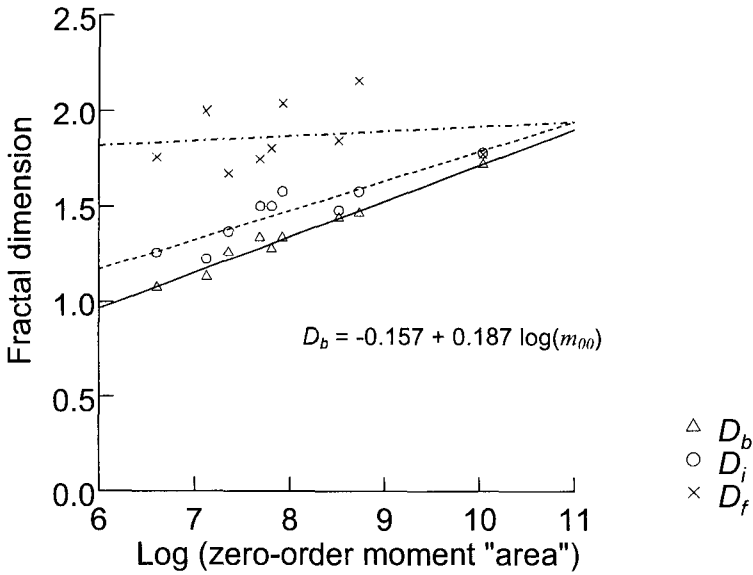

$$D_b = -0.157 + 0.187 \log(m_{00})$$

Figure 2. Relationship between the different fractal dimensions and the logarithm of the zero-order moment which is proportional to the area. Only two regression lines fit the data significantly with d.f. = 7, F = 264, and P=0.000 for D_b and d.f.=7, F=33, and P = 0.001 for D_e.

where ε and θ are respectively the eccentricity and the orientation of the surface distributions [19] and $\mu_{\alpha\beta}$ is defined in the Appendix. The equations sequences on moments are also detailed in the Appendix.

3   Results

Once the probability surfaces have been generated for the different distributions, the fractals and moments were calculated to derive the different descriptors in Table 2.

The fractal dimension values in Table 2 show that the geographical patterns of the distributions of different taxa vary significantly. The high value of D_b (1.719) for taxon 1 indicates that the geographical range of this taxon is more space-filling than all of the other taxa. The correlations are positive between the zero-order moments m_{00}, which represent the area, and both D_b and D_e (Table 3). The two regressions of the logarithm of m_{00} as an independent variable and both D_b and D_e as dependent variables are also significant with a high squared multiple R: 0.974 (Figure 2) for D_b. The correlation is also positive and significant between the shape measured by the eccentricity (ε) and fractal dimension D_b (Table 3).

Table 2. Fractal dimensions and shape parameters or descriptors for the different taxa.

Taxon*	D_b	D_e	D_f	Area(m_{00})	Orientation(θ)	Eccentricity(ε)
Taxon 1	1.719	1.78	1.773	22949	-0.668	20.143
Taxon 3	1.332	1.577	2.039	2761	-0.339	16.221
Taxon 7	1.437	1.479	1.844	5005	-0.451	20.215
Taxon 8	1.276	1.501	1.801	2463	-0.500	16.689
Taxon 9	1.257	1.365	1.67	1565	-0.231	18.064
Taxon 10	1.073	1.255	1.756	740	-0.410	15.023
Taxon 11	1.127	1.222	2.006	1244	0.740	15.886
Taxon 12	1.462	1.576	2.16	6188	0.560	16.337

* Taxa 2, 4, 5 and 6 have only a few records and their distribution maps were not produced.

Table 3. Correlation between the different estimates for the nine taxa derived from their separate distributions.

	D_b	D_e	D_f	Area (m_{00})	θ	ε
Fractal (D_b)	1.000					
Fractal (D_e)	0.930*	1.000				
Fractal (D_f)	0.071	0.108	1.000			
Area (log(m_{00}))	0.987*	0.909*	0.156	1.000		
Orientation(θ)	-0.304	-0.414	0.693	-0.234	1.000	
Eccentricity(ε)	0.759*	0.567	-0.369	0.717*	-0.471	1.000

* Statistically significant correlations

The information fractal dimension D_e concords with the box counting dimension D_b, but the fragmentation dimension D_f differs from both, in that it shows a greater effect from diffusion and fragmentation of the distributions. When $D_f \approx 2$ the distributions are diffuse and highly fragmented (Figure 3 and Table 3). While both taxon 1 and taxon 12
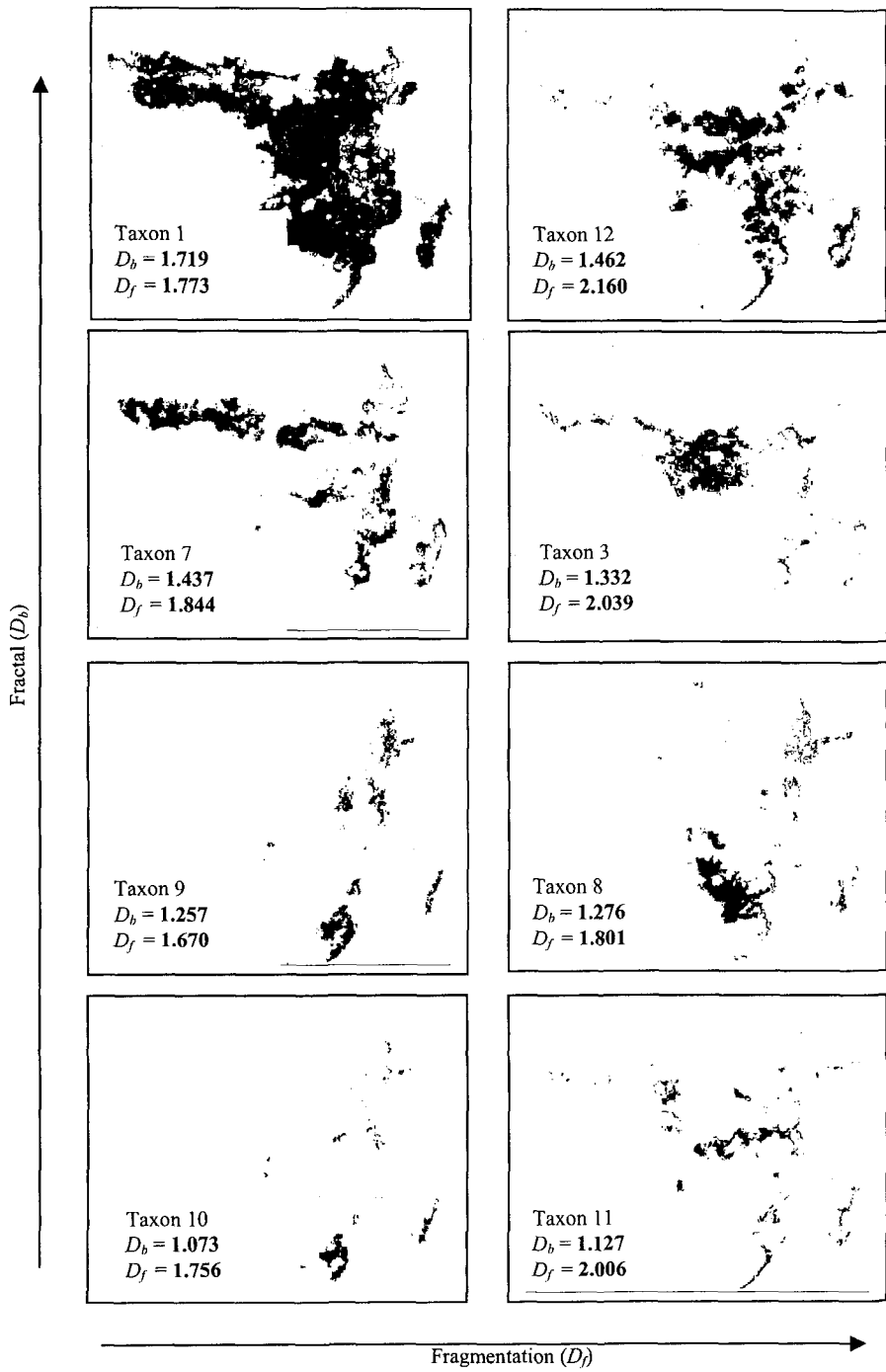
Figure 3. Geographical distribution patterns of the different taxa of wild cowpea. The fractal dimensions show the degree of space-filling and the extent of fragmentation.

have high D_b values they differ significantly in their D_f values. The distribution of taxon 1 is more continuous when compared to that of taxon 12, and the reverse is true for taxa 3 and 11. While both taxon 11 and taxon 3 show the same degree of fragmentation, their space-filling values are different with taxon 3 having relatively more space-filling than taxon 11.

To quantify the amount of information obtained on the distribution patterns using different fractal parameters we have used Shannon's mathematical definition of information, where a robust measure (e.g., D) should reveal values beyond what is expected [20, 21] as per the equation:

$$I(d) = -\sum_D p(D)\log[p(D)] \tag{9}$$

where p_i is estimated as n_i/N, n_i is the number of categories, and N is the total number of individuals recorded in all the categories or classes defined using the mean and the variance of the distribution $p(D)$.

The total variation for each parameter was divided into four categories ($C_1...C_4$; where $\mu - SD \le C_i < \mu + SD$) defined using the mean and the variance. The $I(d)$ was measured for each variable separately. The results in Table 4 show that the box counting method has the highest $I(d)$ value (0.574), while the fragmentation method has the lowest (0.423). Based on $I(d)$ value, the box counting method is more robust in detecting more variability than the information method.

Table 4. The amount of information based on I (index)

Fractal dimension	Mean	SD	$I(d)$
D_b	1.335	0.192	0.574
D_c	1.473	0.173	0.527
D_f	1.866	0.164	0.423

4 Discussion

The present results indicate that fractal geometry can be used to describe plant species' distribution patterns. Fractal dimensions using box counting has been found more discriminating than the other two fractal dimensions based on Shannon index (I) value. Previous work carried out by Maurer [4] has also shown that the box counting method can be useful in the description of species' distribution patterns. Although the fragmentation dimension is not as robust, it remains an important measure of the degree of diffusion and fragmentation of the distributions.

In total these parameters can assist in the quantitative estimation of the geographical pattern of species distribution [4, 22, 23]. Their potential has been recognized in the study of species' patterns of distribution and also in the design and ranking of *in situ* conservation priorities, reserves, and hot spots [4, 8, 10, 22, 24]. Other advantages of fractal geometry are their closeness to model reality, such as in mapping semi-continuous distribution patterns of variation, in contrast to methods that produce binary maps

composed of cells showing only the presence or absence of a species [25]. Fractal algorithms have been used to produce patchy landscapes that remarkably resemble real landscape patterns [23]. The data used in this study will be combined with other data and other predictive methods based on the concept that methods developed in one discipline may by analogy be used to advance other disciplines [26, 27]. The overall aim of the work is to locate where the diversity for the species under study is the highest to help in planning further targeted explorations and in the design of *in situ* conservation sites. In the case of wild cowpea this region is located in the south-eastern part of the African continent.

Appendix

The centroid of the distribution (probability surface) can be expressed in terms of moments as follows:

$$x_c = \frac{m_{10}}{m_{00}} \quad \text{and} \quad y_c = \frac{m_{01}}{m_{00}},$$

where (x_c, y_c) are the coordinates of the centroid. The moments in turn are calculated from the equation below:

$$m_{\alpha\beta} = \sum_x \sum_y x^\alpha y^\beta z_{xy},$$

where the order of the moment is $\alpha+\beta$, x and y are the geographic coordinates, and z_{xy} represents the diversity (number of entities) or probability found in the ith vector [i^{th} square].

The estimates of the transformed latitudinal and longitudinal variances and covariances between them are:

$$S_{\alpha\beta}^2 = \frac{\mu_{\alpha\beta}}{\mu_{00}},$$

where $\mu_{\alpha\beta}$ are the central moments derived from the zero-, first- and second-order moments as per the equation:

$$\mu_{\varepsilon\beta} = \sum_x \sum_y (x - x_c)^\alpha (y - y_c)^\beta z_{xy}.$$

References

1. B. V. Ford-Lloyd and M. T. Jackson, Plant genetic resources: an introduction to their conservation and use. Edward Arnold, London (1986).
2. J. R. Harlan, Crops and Man. American Society of Agronomy, Crop Science Society of America, Madison, Wiconsin (1975).
3. C. Darwin, The origin of species. Bantam Books, NY, USA (1999).

4. B. A. Maurer, Geographical Population Analysis: Tools for the Analysis of Biodiversity. Blackwell Scientific Publications, Alden Press, Oxford, UK (1994).
5. M. E. Ritchie and H. Olff, *Nature* **400**, 557 (1999).
6. B. J. West, Social, Biological and Physical Meta-Mechanisms: A tale of tails. In: Emergent Nature. Novak, M. M. (Ed.), World Scientific Publishing, 77-92 (2002).
7. J. H. Brown, V. K. Gupta B. L. Li, B. T. Milne, C. Restrepo and G. B. West, *Phil. Trans. R. Soc. Lond. B.* **357**, 619 (2002).
8. J. L. Green, Contrasting Self-similarity and Randomness: Species-Area relations in a Californian serpentine grassland. In: Emergent Nature. Novak, M. M. (Ed.), World Scientific Publishing, 113-122 (2002).
9. B.C. Emerson and N. Kolm. Species diversity can drive speciation. *Nature* **434**, 1015 (2005).
10. O. Ovadia, *Cons. Biol.* **17**(5), 1440 (2003).
11. A. Bari, S. Padulosi, G. Ayad, E. Dulloo, T. Hodgkin, A. Martin, J. L. Gonzalez-Andujar, B. Boulouha, L. Sikaoui and A. Brown. Locating and designing *in situ* conservation sites, Conservation and Sustainable Use of Dryland Agrobiodiversity, ICARDA, to appear.
12. S. Padulosi, Genetic diversity, taxonomy and ecogeographic survey of the wild relatives of cowpea [*Vigna unguiculata* (L.) Walpers]. Thèse, Université Catholique, Louvain la Neuve, Belgium (1993).
13. Biodiversity International (2005) http://www.biodiversityhotspots.org/xp/Hotspots
14. P. Jones and A. Gladkov, FloraMap: A Computer Tool for the Distribution of Plants and Other Organisms in the Wild. CIAT, Cali, Colombia (1998)
15. J. Soberon and A. Peterson. Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodiversity Informatics* **2**, 1 (2005)
16. Benoit Fractal Software version 1.3. TruSoft International, Florida, USA.
17. G. P. Nabhan, Wild Phaseolus Ecogeography in the Sierra Madre Occidental, Mexico. *Syst. and Eco. Studies Crop Genepools* **5**, IBPGR (IPGRI), Rome, Italy (1991).
18. A. Bari, A. Della and J. Konopka, Locating diversity using germplasm passport data and herbarium records: Case of *Aegilops* in Cyprus. In: A. Jaradat (Ed.) Triticeae III. Science Publishers, Inc., Enfield, NH, USA (1998).
19. G. W. Awcock and R. Thomas, Applied Image Processing. International Editions, Singapore (1996).
20. G. Battail, Théorie de l'information. Masson, Paris (1997).
21. J. Feldman and M. Singh, *J. Amer. Psych. Assoc.* **112** (1), 243 (2005).
22. S. Hartley and W. E. Kunin, Scale Dependency of Rarity, Extinction Risk, and Conservation Priority, *Cons. Biol.* **17**(6), 1559 (2003).
23. N. Brummitt and E. N. Lughadha, *Cons. Biol.* **17**(5), 1442 (2003).
24. S. M. Pearson and R. H. Gardner, Neutral Models: Useful Tools for Understanding Landscape Patterns. In: J. A. Bissonette (Ed.) Wildlife and Landscape Ecology: Effects of Pattern and Scale. Springer, **C8**, 215-230 (1997)
25. A. Z. Mekjian, *Phys. Rev.* **A 44**, 8361 (1991).
26. S. Hartley and W. E. Kunin, *Cons. Biol.* **17**(6), 1559 (2003).
27. M. R. T. Dale, P. Dixon, M. J. Fortin, P. Legendre, D. E. Myers and M. S. Rosenberg, *Ecography* **25**, 558 (2002).

HIERARCHY OF CELLULAR AUTOMATA IN RELATION TO CONTROL OF CHAOS OR ANTICONTROL

MARIO MARKUS

Max-Planck-Institut für molekulare Physiologie,
Postfach 500247, 44202 Dortmund, Germany
E-mail: markus@mpi-dortmund.mpg.de
Tel: +49-231-97426460

MALTE SCHMICK

Max-Planck-Institut für molekulare Physiologie,
Postfach 500247, 44202 Dortmund, Germany
E-mail: schmick@mpi-dortmund.mpg.de
Tel: +49-231-97426462

ERIC GOLES

Centre for Mathematical Modelling, UMR 2071,
CNRS-Universidad de Chile, Casilla 170-3, Santiago, Chile
E-mail: secrepriv@conicyt.cl
Tel: +56-2-6551389

One-dimensional binary automata are allowed to interact with each other using a 2x2 binary payoff matrix. In this way, chaotic evolution can be controlled or chaos can be induced (anticontrol). For each matrix, a hierarchy of the automata can be established as follows: an automaton with a higher rank controls (or anticontrols) an automaton with a lower rank. For some matrices and some rule number intervals, the ranking depends monotonically on the rule number.

1. Introduction

Nearly one thousand publications are found nowadays in the 'Web of Science' on the subject of chaos control, frequently in connection with cardiological or engineering applications (for reviewing books, see [1,2]). In particular, chaos control has been studied in spatiotemporal systems described by cellular automata (CA) (see e.g. [3] and references therein). In addition, one finds works on 'anticontrol', namely on the induction of chaos in a non-chaotic system. The latter is especially important in the treatment of epileptic seizures, the goal of anticontrol being to desynchronize components of the brain (see e.g. [4]).

297

In the present work we study interactions between CA using 2x2 matrices M. We will show that the possibility of chaos control or anticontrol of one CA by another one leads (for particular matrices M) to well-defined hierarchies of the CA.

The subject of CA hierarchization, parametrization and classification has gone through an intense history in the last twenty years. It starts with the work of Wolfram [5], who studied CAs that operate by iterating a one-dimensional string of cells. Each cell can be in one out of k states (for example: state 0 or state 1, for $k = 2$). The state in a given cell in the next time step is determined (according to a given rule) by the state in that cell and the states of r neighbours at each side of the cell. According to the temporal evolution of the iteration process, Wolfram defines four CA classes:

i) class 1 if evolution leads to homogeneity;
ii) class 2 if evolution leads to periodicity;
iii) class 3 if evolution leads to a chaotic pattern; and
iv) class 4 if evolution leads to complex localized structures, sometimes long-lived.

Langton [6] parametrized the CA defined by Wolfram. For this, he defined a parameter λ equal to the percentage of outputs of the rule which do not lead to zero. It was found [6] that the values of λ roughly correlate with the classes defined by Wolfram. Slightly differing versions of these classes were put into a more formal language in the work of Culik and Yu [7], as well as in Ishii's probabilistic measure theoretic approach [8].

A number of other classifications (alternatives to those mentioned above), some of them involving hierarchization and parametrization, have been reported. Jen [9,10] classified CA by considering the finite sequences which have no pre-image under a given rule. Voorhees and Bradshaw refined that classification by determining some properties of theses sequences [11]. Gutowitz [12] ordered CA as follows: given approximations of the CA by n-step Markov processes, a CA class is the set of CA corresponding to each approximation. Binder [13] then classified the CA according to attractor globality, i.e. according to the property of all initial conditions evolving to the same attractor. This was followed by the work of Kohring [14], who ordered CA according to the number of stable points for a given rule. Markus et al. [15] quantified CA by counting the number of periods larger than 1 resulting from all possible initial conditions for small numbers of cells. Later, Baldwin and Shelah [16] classified CA according to their computational properties (capability to simulate a Turing machine). Then, Dubacq et al. presented an approach based on algorithmic (Kolmogorov) complexity, measuring the "randomness" of evolution through the information content of the local rules [17]. These authors had been inspired by Braga et al. [18], who had investigated how the local rules influence the global behaviour of the CA, but using a topological instead of an algorithmic approach. Recently, Dürr et al. [19] provided yet another approach using Communication Complexity Theory: they partitioned the CA cells into two subsets and determined the information ex-

change for the computation of the CA states that are required by two 'actors', each of them knowing only one of these subsets.

Notwithstanding the large amount of works on CA classification and parametrization, we continued to work on these lines. We did so with practical goals in mind, namely control and anticontrol, which had not been considered before in this context. In order to attain these goals, we let the rules interact with each other. Previously, interaction between rules had ben investigated by Brown [20], who introduced a 'meta rule' that determines the (evolving) rule operating at each cell. Another way of interaction was reported by de Silva et al. [21], who applied stochastically one rule or a second one according to a given probability. CA with rules changing in given subsets of the cells are given by Mori et al. [22]. In the present work, we shall let rules interact via a 2x2 matrix, which is analogous to a payoff matrix in game theory (see e.g. [23] and references therein). We will search for matrices that allow to establish a hierarchy of the CA, considering their abilty to control or anticontrol each other.

2. Algorithm

We consider one-dimensional CA, as introduced by Wolfram [5] with $k=2$ (two states: 0 or 1) and two neighbours ($r=2$), i.e. with five cells in each neighbourhood. We consider totalistic rules, meaning that the output F on any site depends solely on the sum s of states within the neighbourhood. We consider only rules that are called "legal" by Wolfram, i.e those for which the output is 0 for $s = 0$. The code C is given by

$$C = \sum_{s=0}^{5} F(s)\, 2^s \qquad (1)$$

$C=62$ and $C=64$ are not considered, as they trivially lead only to 1's. We are thus left with the 30 rules with $C = 2n$, $0 < n \le 30$, which were graphically displayed by Wolfram in [5]. Wolfram classifies these rules as follows:

  i) class 1 (homogeneous): C=4, 16, 32, 36, 48, 54 and 60 ;
  ii) class 2 (periodical): C=8, 24, 40, 56 and 58;
  iii) class 3 (chaotic): C= 2, 6, 10, 12, 14, 18, 22, 26, 28, 30, 34, 38, 42, 44, 46 and 50;
  iv) class 4: C=20 and 52.

In our algorithm we evaluate the outputs $f_n(t)$ and $g_n(t)$ of two CA at a cell n ($n = 1, 2, 3, ..., N$; cyclical border conditions) at times $t = 0, ..., T$. Considering a 2x2 matrix $M = (a_{ij})$, where $a_{ij} = 0$ or 1, we let the CA interact by writing $f_n(t + 1) = a_{IJ}$ and $g_n(t + 1) = a_{JI}$, where $I = f_n(t)$ and $J = g_n(t)$.

When regarding anticontrol, we say that a rule A has a higher rank (in the CA hierarchy) than a rule B, if the interaction of A and B causes B to become chaotic

or to remain chaotic (class 3). When regarding chaos control, we say that A has a higher rank than B, if the interaction of A and B causes B to become non-chaotic (classes 1,2 or 4 or travelling waves) or to remain non-chaotic. Note that we have defined here class 4 as a controlled state, considering that class 4 automata evolve to homogeneity or to periodicity for finite T. We assign A and B the same rank if both A and B are controlled or anticontrolled.
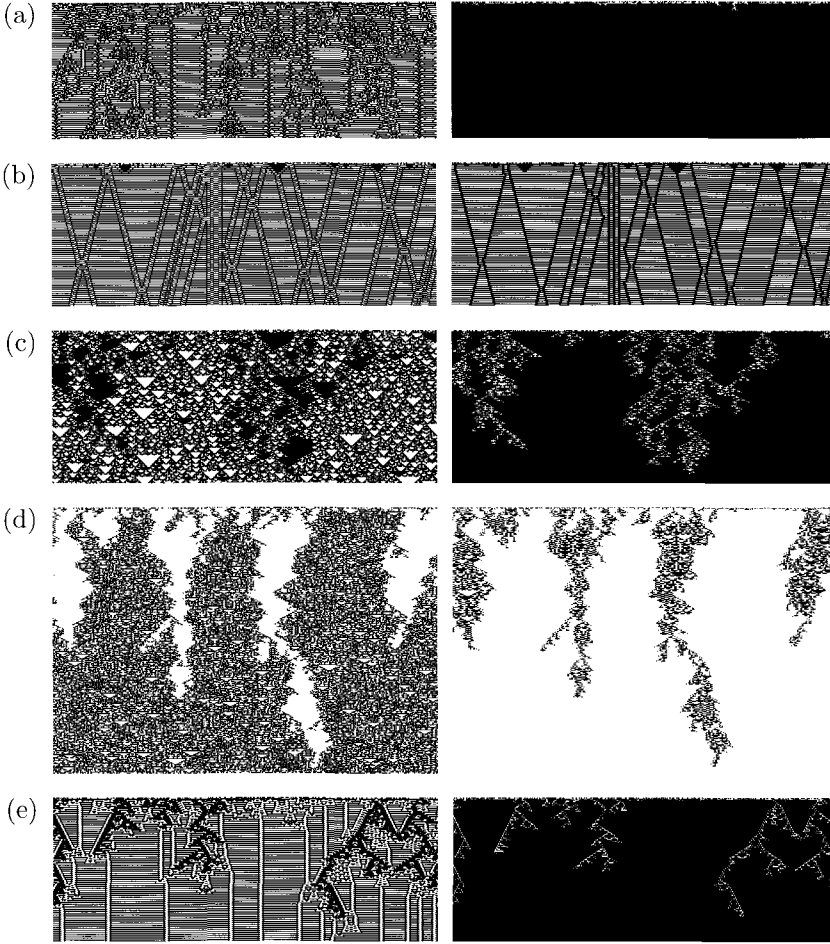

Figure 1. Examples of control. (a) Rule 14 (chaotic) becomes homogeneous (right picture) by interacting via $M = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ with rule 40 (periodical), which itself turns into class 4 (left picture). (b) Rule 34 (chaotic) turns into travelling waves (right picture) by interacting via $M = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ with rule 56 (periodical), which itself also turns into travelling waves (left picture). (c) Rule 46 (chaotic) · becomes homogeneous (right picture) by interacting via $M = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ with rule 44 (chaotic), which itself remains chaotic (left picture). (d) Rule 38 (chaotic) becomes homogeneous (right picture) by interacting via $M = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ with rule 22 (chaotic), which itself remains chaotic (left picture). (e) Rule 28 (chaotic) becomes homogeneous (right picture) by interacting via $M = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ with rule 44 (chaotic), which itself becomes periodical (left picture).

Symmetrical matrices M are not interesting in this work because they drive A and B into the same CA. The matrix $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ is not interesting either, since it leaves both A and B unchanged. $M = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ also changes nothing, except for 0's to become 1's, and viceversa. Thus, there are only four matrices left to investigate: $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

We consider only ranks that are transitive within a given set of rules, i.e. if A (anti)controls B and B (anti)controls C, then A (anti)controls C.

3. Results and discussion

Examples of control and anticontrol with different matrices M are shown in Figs. 1 and 2. In these figures, the evolution of the CA that exerts control or anticontrol, owing to its higher rank, is shown on the left, while the controlled or anticontrolled evolution (lower rank) is shown on the right.

We are unable to recognize any self-consistent ranking when using $M = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. The only facts that can be stated in this case are: rule 32 controls all other rules; rule 30 anticontrols all other rules, except 32.

The results are more satisfactory when using $M = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. In fact, one recognizes a ranking among the chaotic rules, when regarding control. This is illustrated in Fig. 3a. (Note that on Figs. 3a, 3b and 3c the numerical values for the ranking are arbitrary, with a higher rank denoted by a higher vertical position). In the case of Fig. 3a, rule 30 controls all others, while rule 12 is controlled by all others. The ranking is equal in the pairs $(14, 46)$, $(10, 42)$, $(34, 44)$ and $(18, 50)$. The result is less satisfactory for anticontrol: the only non-chaotic CA that can (in a few cases) be anticontrolled are those with $C = 54$ or 58.

The results for $M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ are shown in Fig. 3b. Here, two sets of rules are found for control (2 through 12, and 12 through 50), each set being self-consistent within itself, while there is no well-defined scheme when the interactions occur between rules of the different sets. (Note that the sets are distinguished in the figure by connecting its elements with dashed lines). For anticontrol we found only one set. Note that in Fig. 3b, both for control and anticontrol, one observes something that is not seen in Fig. 3a: monotic dependences (decreasing or increasing) of the ranks with the rule number C.

For $M = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ (Fig. 3c), one obtains one set for control, the ranks of which increase with C, and one set for anticontrol, the ranks decreasing with C. The only exception that disrupts monotonicity is rule 14.

A rough explanation of the monotonicity for control with $M = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ is the following. A CA with relatively low C generates mainly 0's, which are all turned into 1's by M. This swamping with 1's prevents the expansion of chaotic patterns. In the case of anticontrol the argument is opposite: a CA with larger C generates mainly 1's, which are turned into 0's only partially by the action of a CA with a lower C; thus, chaos due to mixing of 0's and 1's appears in the CA with larger C. We are not able to give any explanation going beyond these partial arguments.
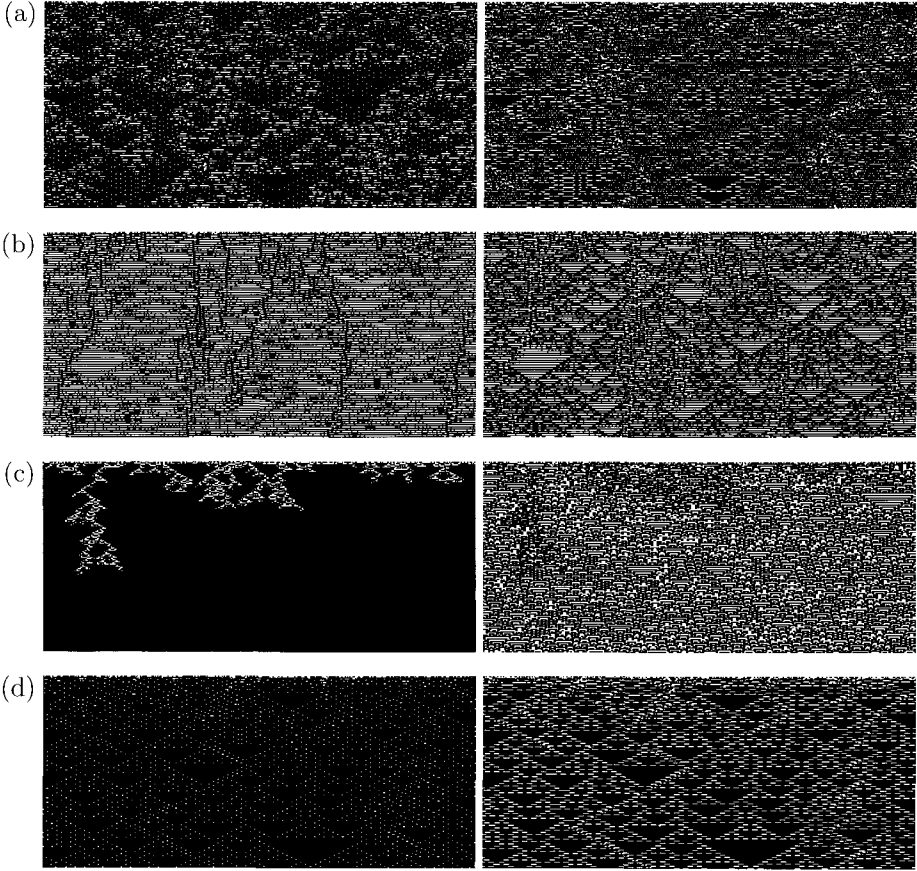
Figure 2. Examples of anticontrol. (a) Rule 16 (homogeneous) becomes chaotic (right picture) by interacting via $M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ with rule 20 (class 4), which itself becomes chaotic (left picture). (b) Rule 36 (homogeneous) becomes chaotic (right picture) by interacting via $M = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ with rule 34 (chaotic), which itself stays chaotic (left picture). (c) Rule 36 (homogeneous) becomes chaotic (right picture) by interacting via $M = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ with rule 24 (periodical), which itself remains homogeneous (left picture). (d) Rule 16 (homogeneous) becomes chaotic (right picture) by interacting via $M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ with rule 24 (periodical), which itself also becomes chaotic (left picture).

4. Conclusion

Hierarchical and parametric classifications of CA have been presented before (see references in the Introduction). The novelty of the present work is that such a classification is performed using control and anticontrol, which are useful operations in many areas (see e.g. [1,2]).

Mathematical proofs, including the monotonicity of the ranks with respect to C, could be a valuable future goal; however such proofs are probably difficult owing to
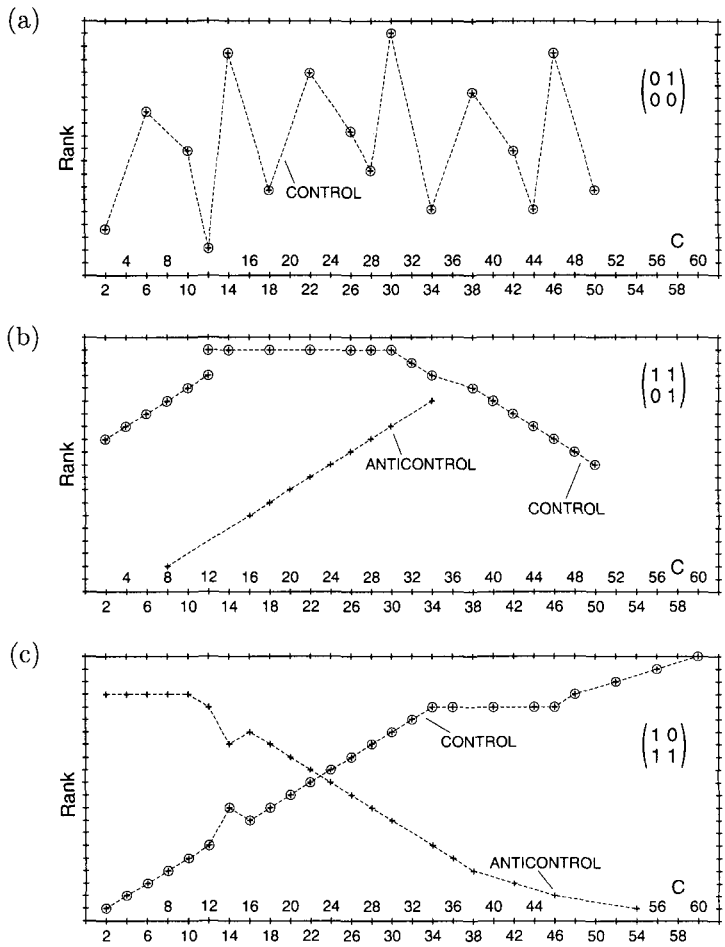
Figure 3. Ranks (ordinates; arbitrary units) of cellular automata rules (codes C in the abscissas) for different interaction matrices M (given on the right of each picture). The following holds within each set of points connected by dashed lines: a rule with a lower rank, i.e. lower abscissa, behaves chaotically (in the cases of anticontrol; crosses) or behaves non-chaotically (in the cases of control; crosses with circles) if it interacts with a rule having a higher rank.

the exceptions (e.g non-monotonicity due to rule 14 in Fig. 3c, as well as several rules which do not fit into the schemes depicted in Fig. 3).

The present work may be considered as a first step in future endeavours concerning automata with physical applications [24,25]. Concretely, this could mean that a parameter is found in such automata, which permits to establish a hierarchy (analogously to our work) for achieving control or anticontrol.

304

References

1. G. Chen and X. Dong, "From Chaos to Order", World Scientific (Singapore, 1998)
2. G. Chen and X. Yu, "Chaos Control", Springer-Verlag (New York, 2003)
3. A. Greilich, M. Markus and E. Goles, Eur. Phys. J. D 33, 279-283 (2005)
4. V.E. Bondarenko, Chaos, Solitons and Fractals 13, 139-154 (2002)
5. S. Wolfram, Physica 10 D, 1-35 (1984)
6. W. Li, N.H. Packard and C.G. Langton, Physica D 45, 77-94 (1990)
7. K. Culik II and S. Yu, Complex systems 2, 177-190 (1988)
8. S. Ishii, Discrete Appl. Math. 39, 125-136 (1992)
9. E. Jen, Complex Systems 1, 1048-1052 (1987)
10. E. Jen, Complex Systems 3, 421-456 (1989)
11. B. Voorhees and S. Bradshaw, Physica D 73, 152-167 (1994)
12. H.A. Gutowitz, Physica D 45, 136-156 (1990)
13. P.-M. Binder, J. Phys. A 24, L31-L34 (1991)
14. G.A. Kohring, Physica A 182, 320-324 (1992)
15. M. Markus, T. Hahn and I. Kusch, Parallel Comp. 23, 1635-1642 (1997)
16. J.T. Baldwin and S. Shelah, Theor. Comp. Sci. 230, 117-129 (2000)
17. J.-C. Dubacq, B. Durand and E. Formenti, Theor. Comp. Sci. 259, 271-285 (2001)
18. G. Braga, G. Cattaneo, P. Flocchini and C. Quaranta Vogliotii, Theor. Comp. Sci. 145, 1-26 (1995)
19. C. Dürr, I. Rapaport and G. Theyssier, Theor. Comp. Sci. 322, 355-368 (2004)
20. D.-B. Brown, Complex Systems 1, 169-180 (1987)
21. L.R. da Silva, H.J. Herrmann and L.S. Lucena, Complex Systems 2, 29-37 (1988)
22. T. Mori, K. Kudo, Y. Namagawa, R. Nakamura, O. Yamakawa, H. Suzuki and T. Uesugi, Physica D 116, 275-282 (1998)
23. H. Fort and S. Viola, Phys. Rev. E 69, 036110 (2004)
24. S. Wolfram, "Theory and Applications of Cellular Automata", World Scientific (Singapore, 1986)
25. D. Stauffer, Lect. Notes in Comp. Sci. 1981, 199-206 (2001)

FRACTAL ANALYSIS OF THE IMAGES USING WAVELET TRANSFORMATION

PETRA JERABKOVA, OLDRICH ZMESKAL, JAN HADERKA

Institute of Physical and Applied Chemistry, Faculty of Chemistry, Brno University of Technology, Purkynova 118, 612 00 Brno, Czech Republic

This article describes a new method of determining fractal dimension and fractal measure using integral (e.g. wavelet – Haar) transformations. The main advantage of the new method is the ability to analyse both, black & white and grey scale (or colour) images. The fractal dimension (box counting fractal dimension) evaluated using the Haar transformation offers in difference to the classic box counting method much wider range of usability. It can be also used to determine the fractal parameters of surfaces or volume of three-dimensional structures e.g. distribution of the mass or electrical charge in the surface or space. The theory of the fractal structures in E-dimensional Euclidean space is described more closely in our previous work.

1   Introduction

Fractal analysis has become one of the fundamental methods of the image analysis. Fractal dimension D and fractal measure K [1] are the two characteristic values, which can be determined by fractal analysis. Fractal measure characterizes the properties of surfaces and edges of image structure, whereas the fractal dimension shows the trends of surfaces and edges changes in relation to the details of structures themselves.

Fractal dimension value lies in the range $D \in \langle 0, E \rangle$, where E stands for Euclidean space dimension ($E = 2$). The limits of the range of the fractal dimension values can be also described as:

- for $D \to 0$ the change of the structure as a function of the size of its measure is maximum (this is characteristic for image structures where analysed structure forms just small part of the complete image, e.g. some colour has very minor representation in the image or the length of the edge between different colours is very short),

- for $D \to 2$ the change of the structure doesn't depend on the change of measure (the analysed structure fills image almost completely, e.g. one colour has majority in the image or the edge fills whole image – Peano curve).

Fractal measure lies in the range $K \in \langle 0, K_{max} \rangle$, where K_{max} is total count of pixels within the image. Using the fractal measure we are able to determine how much of the image is covered by the analysed structure. For example fractal measure of Peano curve is equal to the total count of the pixels within the image (or we can say that the coverage by the edge is 100%, $K/K_{max} = 1$ respectively). The fractal dimension of the edge between black and white within the image is then $D_{BW} = 2$.

The method most often used to determine the fractal parameters of the image is called box counting method. This method allows analysing of black & white pictures created by changing of threshold value of colourful image structure by specified criteria

(e.g. on RGB, HSV, HSB space) – the thresholding. The results of fractal analysis (using the box counting method) are box counting fractal dimension and box counting fractal measure.

This article describes a new method of determining fractal dimension and fractal measure using integral (wavelet – Haar) transformation. The main advantage of the new method is the ability to analyse grey scale (or colour) images without lengthy thresholding and calculating the box counting fractal dimension D and fractal measure K for each thresholded image in the set.

More detailed results of the analysis using this new method can be found on our web pages http://www.fch.vutbr.cz/lectures/imagesci. The interpretation of the results obtained from the black & white image analysis (using the new method) remains the same as from the box counting method of black & white pictures. The main disadvantage of the new method may be seen in smaller set of data used for regression analysis as in box counting method. There can be obtained only 10 independent values per image with 1024×1024 pixels matrix that can be used to determine the fractal dimension D and fractal measure K. Such disadvantage can be turned into an advantage as removing redundant use of pixels minimizes the error introduced by reusing the values of wrongly determined values of pixels.

2 Box counting method

Analysis of black & white images is usually performed by box counting method [2]. *Figure 1* shows how is the method used. A net of different size ($1 \times 1, 2 \times 2, 3 \times 3, 4 \times 4$, etc. pixels) is placed on the analysed image and the total count of black boxes N_B, white boxes N_W and black&white N_{BW} boxes is determined.
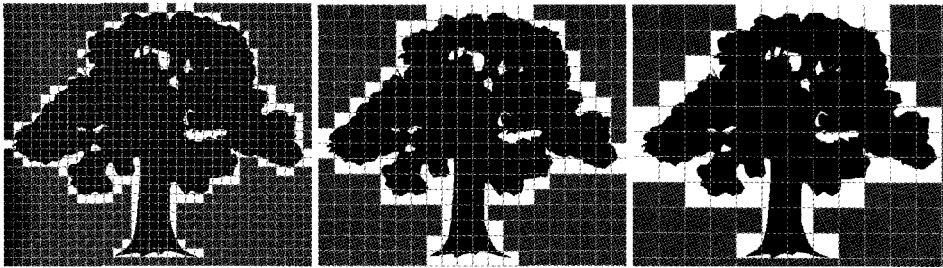
Figure 1. Determination of the fractal measure and fractal dimension using the box counting method.

The function of the count of pixels on the net size $\varepsilon = 1/n$, where ε is the size of pixels and n is the scale, is used to determine three different types of fractal dimension: black area, white area and their edge (D_{BBW}, D_{WBW}, D_{BW}). To calculate the dimension D_{BBW} and D_{WBW}, the *count* of black and black & white (BW) $N_{BBW} = N_B + N_{BW}$ or white and black & white (BW) $N_{WBW} = N_W + N_{BW}$ boxes is utilised. The variation of fractal dimension with the count of boxes can be written

$$N(\varepsilon) = K \cdot \varepsilon^{-D} , \qquad\qquad N(n) = K \cdot n^{D} , \qquad\qquad (1)$$

respectively, where K is the fractal measure and D is the fractal dimension (without their indexes). From these terms the fractal dimension is given by the slope of function (1)

$$D = -\frac{\mathrm{d} \ln N(\varepsilon)}{\mathrm{d} \ln \varepsilon} = \frac{\mathrm{d} \ln N(n)}{\mathrm{d} \ln n} \cdot \qquad\qquad (2)$$

The area filled by the examined set (points) can be expressed as

$$S(n) = \frac{N(n)}{n^2} = K \cdot n^{D-2} \qquad\qquad (3)$$

and the length of lines in figure as

$$L(n) = \frac{N(n)}{n} = K \cdot n^{D-1} \qquad\qquad (4)$$

respectively.

The *Figure 2* shows the dependency obtained from the model image of a tree for black & white squares. The image shows that the function is almost linear (structure of the image is fractal). There is clearly a visible error for small sizes of the boxes (introduced by the interpolation of values) and also error for the very big sizes of the boxes (introduced by saturation of the number of boxes).
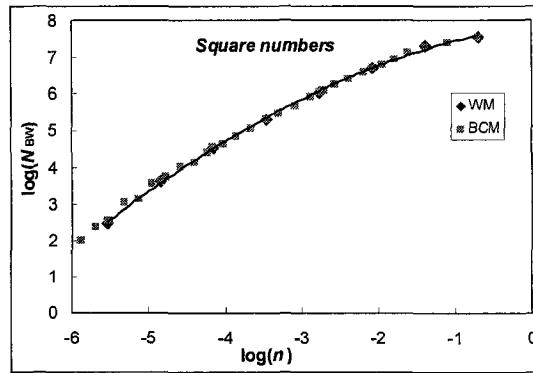
Figure 2. Determination of the fractal measure and fractal dimension using the box counting (see Figure 1) and for wavelet method (see Figure 3 - Figure 5).

Acquired fractal measures K_{BBW}, K_{WBW}, K_{BW} can be used for determination of the black $- S_B$ (white $- S_W$ respectively) areas of threshold figure

$$S_{\mathrm{B}} = \frac{K_{\mathrm{BBW}} - K_{\mathrm{BW}}}{K_{\mathrm{BBW}} + K_{\mathrm{WBW}} - K_{\mathrm{BW}}} , \quad S_{\mathrm{W}} = \frac{K_{\mathrm{WBW}} - K_{\mathrm{BW}}}{K_{\mathrm{BBW}} + K_{\mathrm{WBW}} - K_{\mathrm{BW}}} \qquad\qquad (5)$$

and the length of divided lines L of threshold figure

$$L = \frac{K_{BW}}{K_{BBW} + K_{WBW} - K_{BW}} \cdot \tag{6}$$

The results are taken in relative units (multiplied by 100 in percent). Results in pixels can be taken after multiplying this number by area size of the figure (K_{max}).

3 Wavelet method

The description of the box counting method shows that it is variation of integral methods. Using that knowledge the authors tried to find the best integral transformation that can be used to evaluate the fractal parameters of image structures. The driving force of our effort was to simplify the algorithm for determination of the fractal dimension and to extend the range of possible interpretation of the results. It was found that already very simple wavelet transformation – Haar transformation [3] gives us the *identical results* of values of fractal parameters (for threshold figure) as the box counting method.

Haar transformation (HT) is linear orthogonal transformation with signum function (rectangle function) base. The HT transforms real image $f(m, n)$ into discreet spectrum represented by real function $F(k, l)$

$$F(k,l) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f(m,n) h_{m,k} h_{n,l}, \tag{7}$$

where $h_{n,k}$ and $h_{m,l}$ are the coefficients of so called Haar matrix H_N

$$H_0 = 1, \quad H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \cdots . \tag{8}$$

Haar matrix H_n is the order of 2^n (2^n dimensional matrix).

$$2 \times 2 \qquad\qquad 4 \times 4 \qquad\qquad 8 \times 8$$
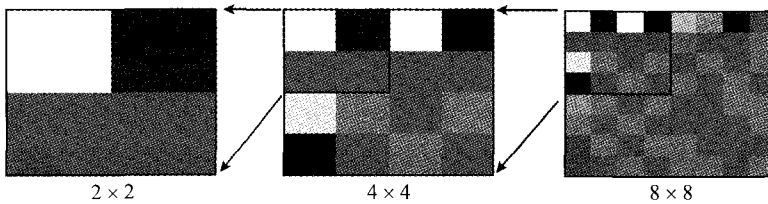
Figure 3. Spectra of image structure determined by Haar transformation of image from Figure 1 (wavelet method)

The method requires (like the Fast Fourier Transformation) a square matrix of data given by square of two. *Figure 3* show the resulting spectra and images of the tree (*Figure 1*) of size 1024×1024 pixels. The spectra are calculated using the matrix of Haar transformation (transformation base) of size (matrix and image) 1024×1024

pixels. *Figure 3* shows top left corner cut from the complete matrix (2×2, 4×4, 8×8 pixels). The top left item of the matrix stores information about amount of black pixels within the image. The information about count of black pixels in each quarter of the image can be then determined from the matrix obtained by the reverse Haar transformation (such matrix is then half of the size of the original matrix), see *Figure 4*.

The *Figure 5* shows information about count of black pixels in the quarter of the matrix. Shades of grey on *Figure 4* are equivalent to the count of black & white (BW) pixels in the quarter. Fractal dimensions of black, white and black & white can be determined as it was possible when using box counting method (D_{BBW}, D_{WBW}, D_{BW}). The parameters obtained by our method are identical to parameters determined by the box counting method for matrixes of $2^n \times 2^n$ (1×1, 2×2, 4×4, 8×8, etc.) pixels.

| 64 × 64 | 32 × 32 | 16 × 16 |
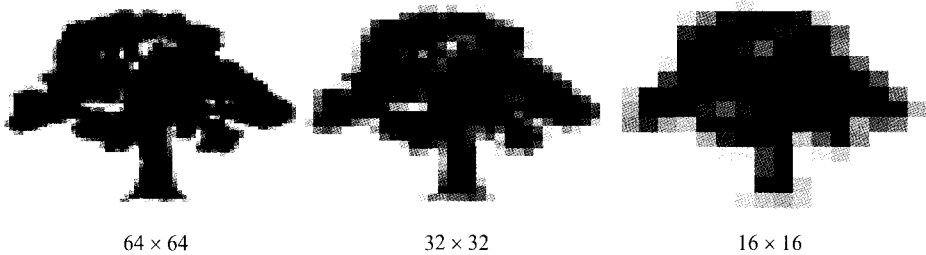
Figure 4. Determination of the fractal measure and fractal dimension using wavelet method (grey scale squares)

| 64 × 64 | 32 × 32 | 16 × 16 |
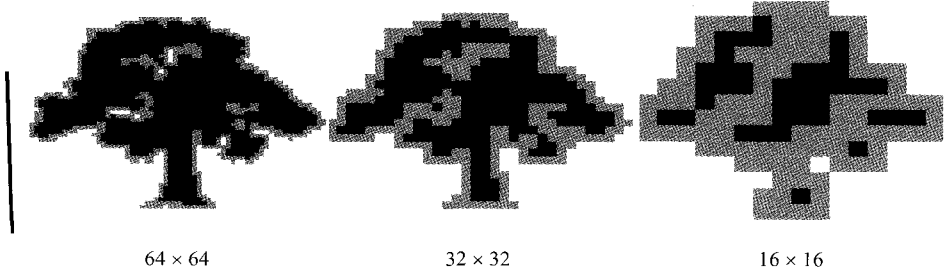
Figure 5. Determination of the fractal measure and fractal dimension using wavelet method (tresholded squares)

The function of count of black & white squares is displayed in *Figure 2* (these determine the characteristic of the interface between black and white areas in the image). The fractal dimension determined by the box counting method was $D_{BW} = 1.1053$. The graph displayed shows that the function is not linear. The exact derivative of the function is shown in *Figure 2* right. Error in the big squares area (on the right side of the graph) is caused by error of applying grid on the image (error is approximately same for both methods). Error in the small squares area (left side of the graph) is caused by wrong interpolation of the data. Inflection point of the graph curve is exactly at the point equal to the value of the fractal dimension of the structure.

The acquired fractal measures K_{BBW}, K_{WBW}, K_{BW} can be used for determination of black and white areas of figure and the dividing lines lengths using the same equations as for box counting method, see Eqs. (5) and (6).

4   Algorithm of fractal parameters calculation using the Haar transformation

The method of fractal parameters calculation described in the previous chapter is relatively time-consuming. It requires performing the direct Haar transformation of analysed picture (with size $2^n \times 2^n$ pixels) and $(n-1)$ reverse Haar transformations of spectrums, which were previously filtered by low-pass filtering $(n-1) \times (n-1)$, $(n-2) \times (n-2)$, ... pixels. The fractal parameters are then calculated using standard technique (like box-counting method) from the sequence of the obtained pictures with flattened details.

It was found by detail problem analysis that the algorithm above can be simplified by simple adding of four adjacent pixels (which have values 0 and 1 in the original thresholded picture). After the first step we obtain picture of size $2^{n-1} \times 2^{n-1}$ pixels with values from 0 to 4, after the second step the picture will have its size $2^{n-2} \times 2^{n-2}$ pixels with values from 0 to 16, etc. The N_B of the picture is then the number of pixels with value 0, the N_W is number of pixels with maximal value (i.e. 1, 4, 16, ... , 2^{2n}) and the N_{BW} is the number of the rest of pixels. The log N = f(log n) dependency allows to determine the fractal dimension and fractal measure of the tresholded picture using the standard techniques. The analysis of a black box gives fractal dimension $D_W = 0$ (number of white pixels is always zero) whereas the analysis of a box containing only white pixels results in $D_W = 2$.

5   Analysis of three-dimensional structures

In the case of analysing a grey scale picture, it is supposable, that the shade represents a profile of a three-dimensional structure (see e.g. *Figure 6*). In this case the particular pixels have values from 0 (black) to 255 (white). Adding of values of foursome pixels results in picture of $2^{n-1} \times 2^{n-1}$ pixels size with values from 0 to 1020. After the second iteration a $2^{n-2} \times 2^{n-2}$ pixels sized picture with values from 0 to 2040 is obtained, etc.
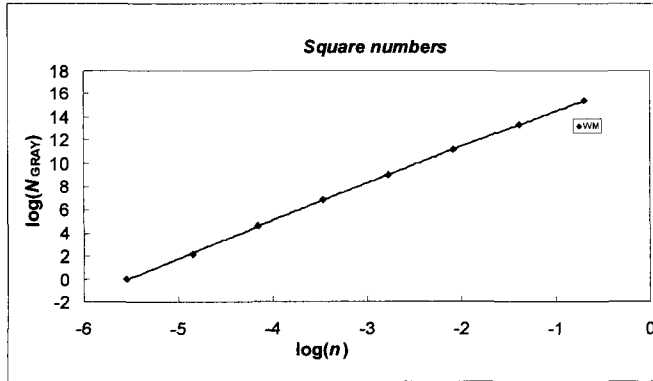

2D projection   3D projection
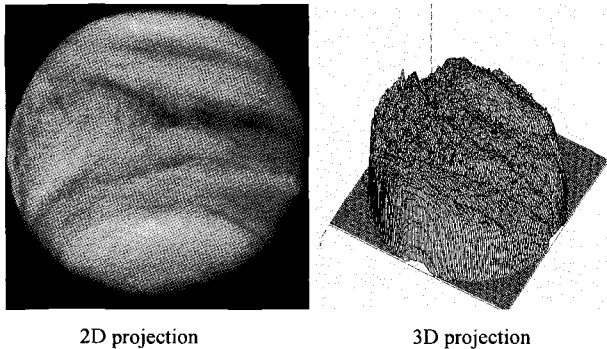
Figure 6. Determination of the fractal measure and fractal dimension of 3D structure (fractal dimension of surface) using wavelet method

The N_B of each individual picture is then the number of pixels with value 0, the N_W is the number of pixels with maximal value (i.e. 1020, 4080, ... , $255 \cdot 2^{2n}$) and the N_{BW} is

the number of the remaining of pixels. The log $N = f(\log n)$ dependency allows to determine the fractal dimension and fractal measure of the tresholded picture using the standard techniques. The analysis of a black box gives fractal dimension $D_W = 0$ (number of white pixels is always zero) whereas the analysis of a box containing only white pixels results in $D_W = 3$.

The volume filled by the examined set (points) can be expressed as

$$V(n) = \frac{N(n)}{n^3} = K \cdot n^{D-3} . \tag{9}$$

The following equations determines the areas under (V_B) and above (V_W) of the 3D object surface

$$V_B = \frac{K_{BBW} - K_{BW}}{K_{BBW} + K_{WBW} - K_{BW}}, \quad V_W = \frac{K_{WBW} - K_{BW}}{K_{BBW} + K_{WBW} - K_{BW}} \tag{10}$$

and area size S

$$S = \frac{K_{BW}}{K_{BBW} + K_{WBW} - K_{BW}} \tag{11}$$

in relative units (in percent respectively). Volumes and areas in volume pixels (voxels) can be obtained by multiplying the results by the maximal volume. It depends on the area size of figure and its colour depth (e.g. for eight bit depth is given by multiplying by the number $2^8 - 1 = 255$). Coefficient which determines how many times is the fractal surface area higher then the horizontal surface area (size of figure) can be also calculated.

Figure 7. Determination of the fractal measure and fractal dimension of 3D structure (fractal dimension of surface) using wavelet method

Fractal dimension of surface of a fractal structure in *Figure 6* is $D_{BW} = 2.263$. Fractal dimension of the structure volume was found to be $D_{BBW} = 2.877$ and fractal dimension of the surrounding volume is $D_{WBW} = 2.900$. Volumes of the object and surroundings have the fractal dimension near the value of the Euclidean dimension of volume $E = 3$.

This implies that the object has nearly the Euclidean shape. Fractal dimension of the surface is significantly bigger than the Euclidean dimension of plane $E = 2$ and has therefore evidently fractal nature. The fractal measure of the interface determines the size of the fractal structure surface (in pixels) and is $K_{BW}/K_{max} = 362.79\%$ of the image's planar surface, where $K_{max} = (512 \times 512)$ pixels. Fractal measure of the structure volume is $K_{BBW}/K_{max} = 40.97\%$ and its surroundings $K_{WBW}/K_{max} = 59.03\%$, where $K_{max} = (512 \times 512 \times 255)$ voxels (volume pixels). The fractal measure of the interface tells us therefore that the surface of the fractal structure is more than three times larger than the surface of plane.

6   Conclusion

It is worth mentioning that the fractal dimension of the deterministic fractal structures (e.g. Sierpinsky carpet) is immutable in the whole range of sizes of the net applied on the image using box counting method (or level of filter using wavelet – Haar – transformation) and is the same as the theoretical fractal dimension value. Similar fractal dimension values were also obtained using other types of integral, e.g. the Fourier transformations. In this case the values of fractal parameters are influenced by broader vicinity (step function is replaced by harmonic function).

The fractal dimension evaluated using the Haar transformation offers in contrast to the classic box counting method a much wider range of usability. The method can also be used to determine the fractal dimension of colour structures (e.g. greyscale or full colour images). This also allows determining the fractal dimensions of surfaces specified by shades of grey or colour components. This can be also used to determine the fractal parameters of surfaces or three-dimensional structures like distribution of the mass in the space or electrical potential in space. The theory of the fractal structures in E-dimensional Euclidean space is described more closely e.g. in [4, 5].

Software for fractal analysis of the images is provided free on the Internet http://www.fch.vutbr.cz/lectures/imagesci.

Literature

1.   B. B. Mandelbrot, Fractal geometry of nature. New York: W.H. Freeman and Co. (1983)
2.   M. J. Barnsley, Fractals everywhere, New York:Academic Press Inc. (1993)
3.   G. Strang, Wavelets Transforms versus Fourier Transforms, Bulletin of the American Mathematical Society, 28, 288 (1993)
4.   O. Zmeskal, M. Nezadal, M. Buchnicek, Chaos, Solitons & Fractals, 17, 113 (2003)
5.   O. Zmeskal, M. Nezadal, M. Buchnicek, Chaos, Solitons & Fractals, 19, 1013 (2004)

CLUSTERING PHENOMENA IN THE TIME DISTRIBUTION OF LIGHTNING

LUCIANO TELESCA

Institute of Methodologies for Environmental Analysis, C.da S.Loja, 85050 Tito (PZ), Italy

MARINA BERNARDI, CINZIA ROVELLI

CESI, Via Ribattino 54, 20134 Milano, Italy

Two years of lightning data, measured during 2002 and 2003 in two areas of Italy, Lombardia (northern Italy) and Campania (southern Italy), have been investigated in order to reveal clustering properties in their time dynamics. The Allan Factor analysis has been performed. Our findings point out to the existence of time-clustering in the sequence of lightning time-occurrences, characterized by the presence of two scaling regions, which reveal intra-cluster (inside individual thunderstorms) and inter-cluster (among successive thunderstorms) time-clustering properties of lightning data. A crossover timescale is identified, which could be connected with the mean duration of a thunderstorm.

1   Introduction

Lightning, defined as a high and transitory electric current, are usually generated by electrostatic charge separation occurring in thunderstorm clouds [1]. The most accepted explanation for this process is the friction and collision among ice and water particles traveling in convective drafts in the cloud. Charged particles tend to migrate to far regions of the cloud, typically the top and bottom, to reach a configuration of a dipole or multi-pole; earth polarizes opposite to the bottom cloud charge becoming another pole. The electric field intensity between dipoles can reach the breakdown level and originate a lightning discharge [2].

The total discharge is called a "flash" and is a complex process involving one or more subsequent current flows; each of these currents is called a "stroke". A flash lasts for about half a second, while each stroke takes around one millisecond [3].

A flash can thus happen between two clouds or into the same cloud or from cloud to ground. This last category, even if the less numerous, is by far the more known and studied [4]. Cloud-to-ground flashes could be subdivided in positive and negative from the current polarity or in upward and downward from the direction of motion. Positive cloud-to-ground discharges are rare, around 10% of the cloud-to-ground total; they can transfer to ground a considerable quantity of charge, compared to the negative lightning, and can have higher current intensity than negative ones. Usually positive flashes are much more common in winter thunderstorms [5].

Thunderstorms vary from region to region on the Earth, due to different latitude, meteorological conditions, orography [6]. A mean of 100 s-1 lightning strikes the earth, of which much more fall over land than on the oceans; tropical thunderstorms present a

higher number of flashes [7]. In our case the region under study is in the temperate area, in southern Italy. In Italy the thunderstorms are of three main kinds:

1. large fronts: these are formed by collision between incoming cold air mass and pre-existent hot air mass. These thunderstorms are joined to atmospheric disturbance and they are dynamic thunderstorms. In Italy large front thunderstorms come generally from North-West;

2. local convective cells: these are produced by convective cells that develop usually during afternoon on mountains and during late-afternoon or evening on plain, when the unstable weather conditions are more favorable. This thunderstorm kind is typical in summer season;

3. orographyc thunderstorms: they are generated by raising air mass due to the presence of a slope. They develop typically in Alpine and Appenninic regions.
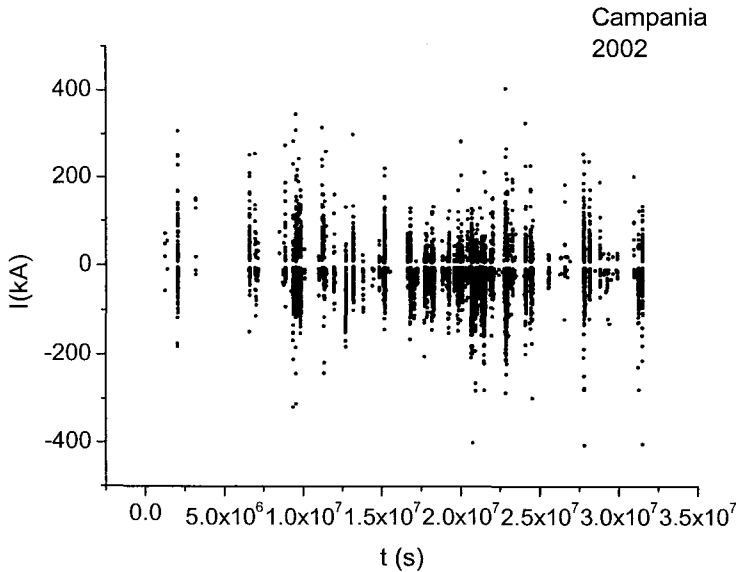
Figure 1. Time distribution of lightning in Campania (southern Italy) and Lombardia (northern Italy) in 2002 and 2003.
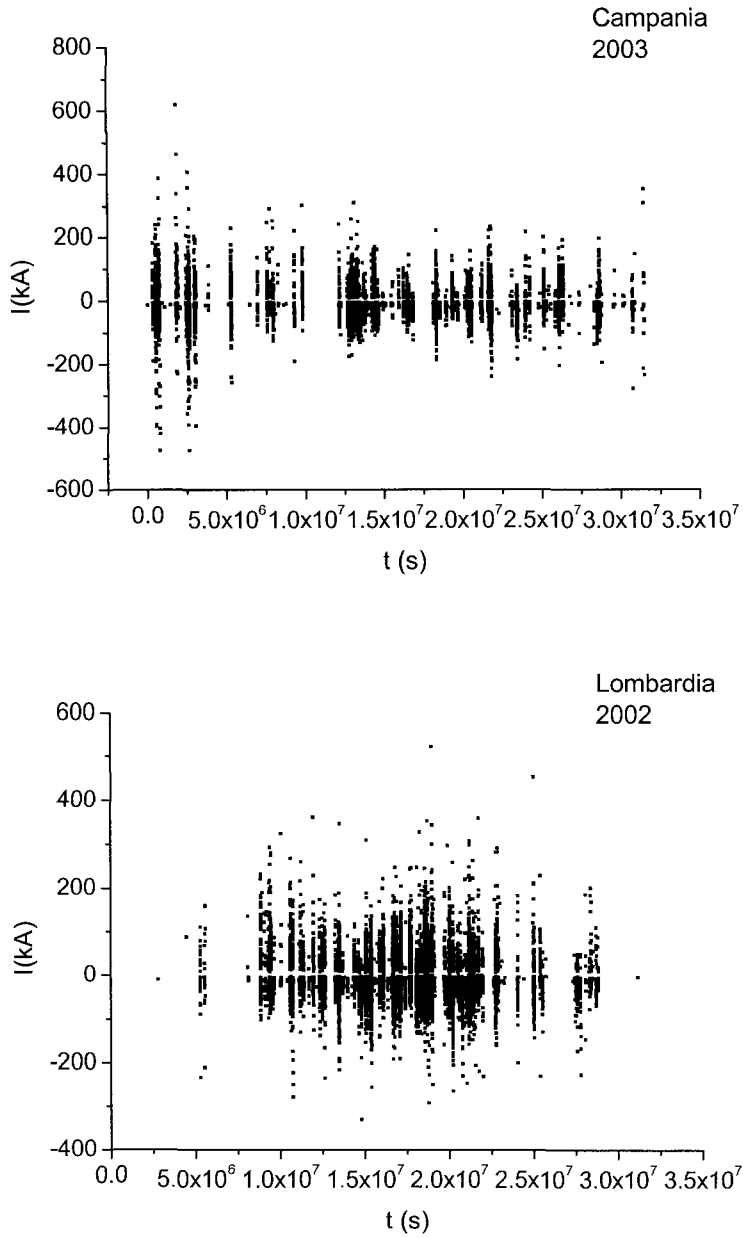
Figure 1 (cont.).
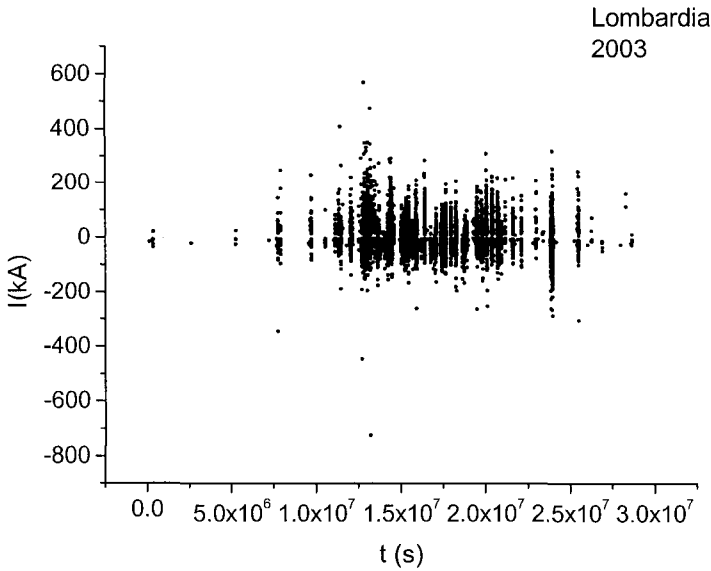
Figure 1 (cont.).

2 Lightning Data

In this study we analyze the temporal properties of the cloud-to-ground lightning activity. The data have been provided by the lightning location systems (LLS) of Sistema Italiano Rilevamento Fulmini (SIRF) [8], which is a network detecting cloud-to-ground lightning activity over all of Italy, composed by 16 IMPACT (Improved Accuracy from Combined Technology) sensors [9] located throughout Italian territory. The mean accuracy of the system is 500 m for both co-ordinates of the point of impacts. SIRF allows the detection of single strokes with peak current value greater than about 2.5 kA. The nominal flash detection efficiency is roughly 90% over all the area [8]. The SIRF system provides a set of information for each cloud-to-ground lightning: date, time of occurrence, lightning impact coordinates, peak current, stroke number, polarity, statistic parameters.

A sequence of 34,502 (during 2002) and 29,200 (during 2003) lightning flashes was recorded in Campania (southern Italy), and 55,496 (during 2002) 48,540 (during 2003) occurred in Lombardia (northern Italy). Figure 1 shows the time distribution of the lightning. The association of lightning into clusters (time-clusterization) is rather clear by visual inspection from Figure 1.

3   Allan Factor Analysis

The aim of the present study is the identification and quantification of the clustering in lightning sequences. The power spectral density (PSD) S(f), obtained by means of a Fourier Transform (FT) of the series, was the standard technique used to investigate the temporal properties of a time series. The PSD gives information on how the power of the series is concentrated at various frequency bands [11]. This information allows identifying periodic, multi-periodic or non-periodic frequency patterns. Usually the logarithmic power spectrum plot, that is the power spectrum plotted in log-log scales, is used to analyze broadband behaviour. The power-law dependence (linear on a log-log plot) of the PSD, given by S(f)~ f^{α}, is a hallmark of the presence of time-scaling in the data. The properties of the series can be further classified in terms of the numerical value of the spectral exponent $\alpha=0$ or $\alpha\neq0$ feature white noise or pink noise time series, thus indicating the absence or presence of time correlations respectively [12].

A lightning sequence is a temporal point process, that is a set of events occurring at random locations in time [13], and the simple application of the FT is not possible. We can represent the lightning data by a finite sum of Dirac's delta functions centered on the occurrence times t_i, with amplitude A_i proportional to the current:

$$y(t) = \sum_{i=1}^{N} A_i \delta(t - t_i)$$, (1)

where N represents the number of events recorded.

Then dividing the time axis into equally spaced contiguous counting windows of duration τ, which is called timescale, we produce a sequence of counts $\{N_k(\tau)\}$, with $N_k(\tau)$ denoting the number of flashes in the k-th window [14]:

$$N_k(\tau) = \int_{t_{k-1}}^{t_k} \sum_{j=1}^{n} \delta(t - t_j) dt$$. (2)

where t is the integration variable. We can define the following quantity, the Allan Factor (AF),

$$AF(\tau) = \frac{< (N_{k+1}(\tau) - N_k(\tau))^2 >}{2 < N_k(\tau) >}$$, (3)

which is related to the variability of successive counts [15, 16], where <..> indicates the expectation value. The AF has been largely used to investigate the time dynamics of earthquakes [17].
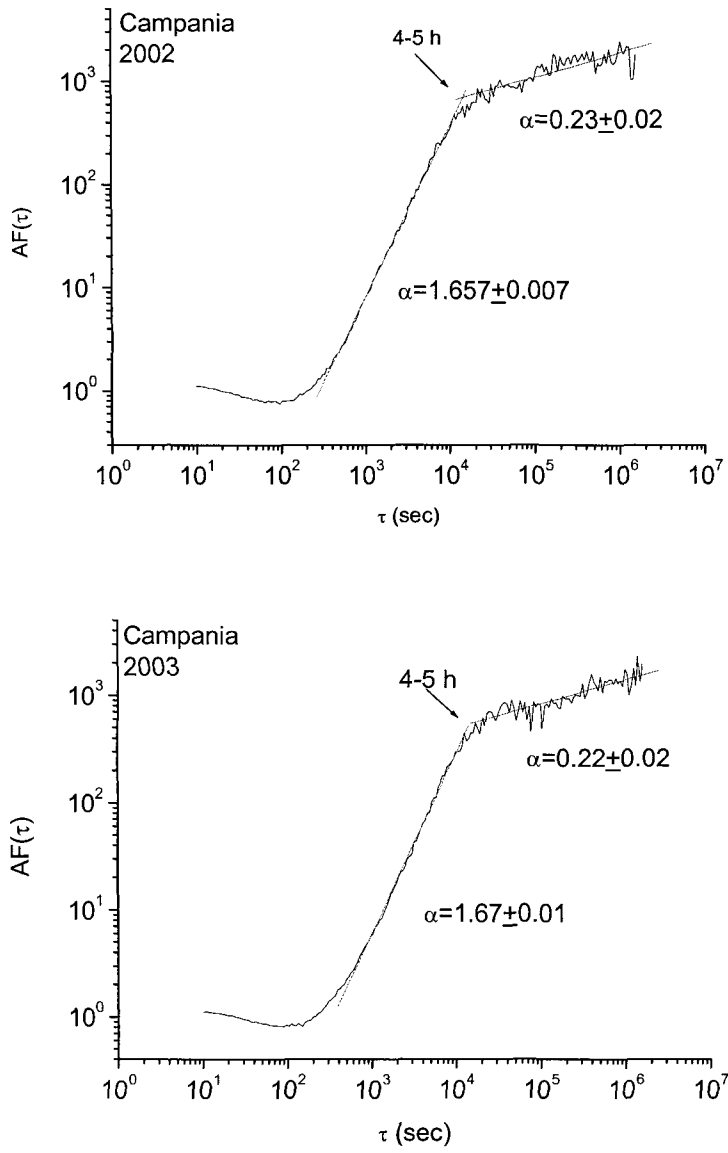
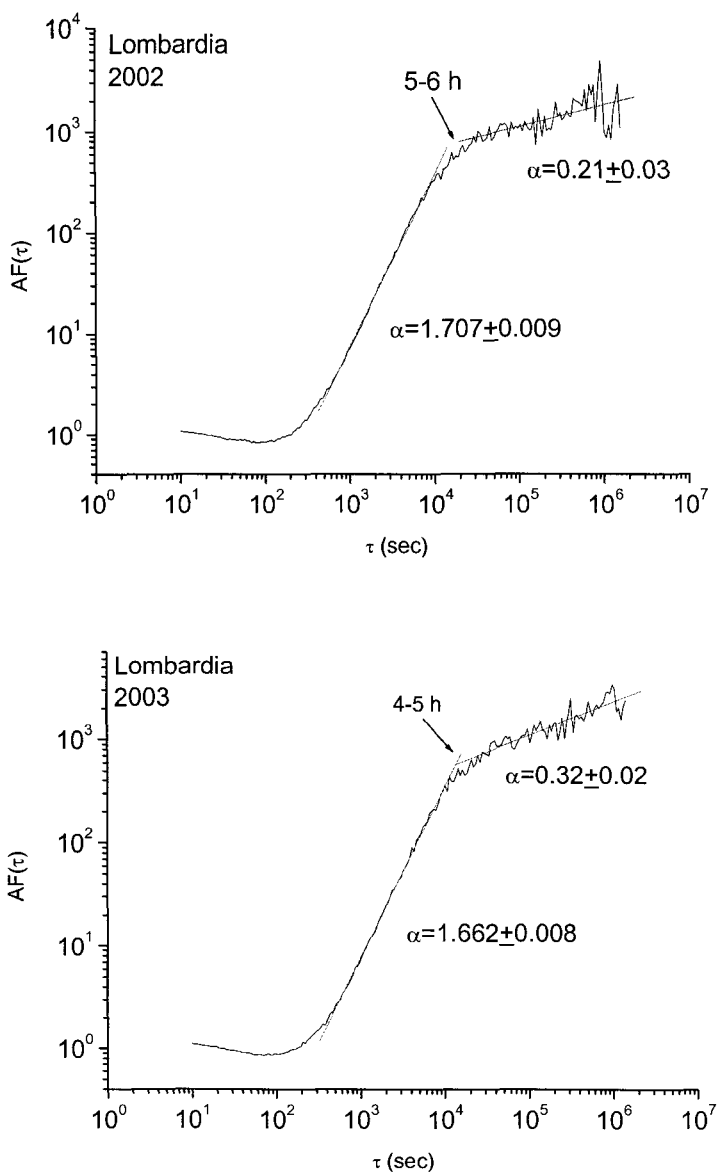Figure 2. Allan Factor curves for the lightning data plotted in Figure 1.

Figure 2 (cont).

If the point process is time-clusterized, then the AF varies with the timescale τ with a power-law form:

$$AF(\tau) = 1 + \left(\frac{\tau}{\tau_1}\right)^{\alpha}$$

(4)

over a large range of timescales τ, with the exponent α, which quantifies the strength of time-clusterization; τ_1 is the fractal onset time and marks the lower limit for significant scaling behavior in the AF, so that for $\tau \ll \tau_1$ the time-scaling property becomes negligible at these time scales [18]. AF assumes values near unity for Poisson processes [19]. Therefore, if $\alpha=0$ the point process is Poissonian, which means that the series is memoryless and constituted by independent events; while if $\alpha>0$ the process is characterized by time-scaling behaviour, which means that the series is time-correlated.

4   Results

Figure 2 shows the relation AF$(\tau)\sim\tau$ in log-log scales for the analyzed ligthning sequences. Eq. 3 has been calculated for timescales τ ranging between 10 s and T/10, where T is the total observation period. The first striking feature is that the lightning activity is not globally Poissonian, since the AF curves are not flat for all the timescales τ, except for the lower timescales, where the behaviour can be considered Poisson-like. In particular, two distinct scaling regions are visible in all the AF curves, limited by two crossover timescales, $\tau_1\approx400$-500 s and $\tau_2\approx4$-6 hours. In the first region, ranging between τ_1 and τ_2, the AFs scale with τ with scaling exponent $\alpha\approx1.7$. The second scaling region is characterized by the scaling exponent $\alpha\approx0.2$-0.3. The crossover timescale $\tau_2\approx4$-6 hours can be linked with the mean duration of a thunderstorm; therefore, the first scaling region describes the time-clustering properties of the lightning activity inside a thunderstorm (intra-cluster correlations), while the second scaling region describes those among different thunderstorms (inter-cluster correlations). From the numerical value of α we deduce that the time-clustering of lightning inside a thunderstorm is much stronger than that evaluated for lightning charged to the ground during different storms.

5   Conclusions

This study shows that lightning activity is characterized by time-clustering behaviour. Two years of lightning data recorded in two areas of Italy have been anlyzed. The AF methodology has disclosed the presence of two different time-scaling regimes, connected with the time-clustering behavior of lightning inside individual thunderstorms (intra-cluster) and that among successive thunderstorms (inter-cluster); the crossover timescale between the two scaling regions could be connected with the average duration of a thunderstorm. This study proposes a new approach in investigating the time dynamics of lightning sequence.

Acknowledgments

References

1. R. B. Anderson and A. J. Eriksson, *Electra* **69**, 65 (1980); K. Berger, *Elektrotechnik* **82**, 249 (1967); E. P. Krider, *J. Geophys. Res.* **94**, 13145 (1989); E. P. Krider and R. J. Blakeslee, *J. Electrostat.* **16**, 369 (1985).
2. B. F. J. Schonland, *Proc. R. Soc. London* **A118**, 233 (1928); M. A. Uman, The Lightning Discharge (Academic Press, Orlando, Florida, 1987); B. Vonnegut, The physics of Thunderclouds, in Handbook of Atmospherics, ed. H. Volland, (CRC Press, Boca Raton, Florida 1982); B. Vonnegut, *Bull. Am. Meteor. Soc.* **34**, 378 (1953).
3. V. A. Rakov and M. A. Uman, *J. Geophys. Res.* **95**, 16561 (1990).
4. R. E. Orville, *Weatherwise* **25**, 108 (1972).
5. R. L. Holle and A. I. Watson, *Wea. Forecast.* **11**, 600 (1996); K. Miyake and T. Suzuki, *IEEE Trans. Pow. Del.* **5**, 1370 (1990).
6. V. Rakov and M. A. Uman, Lightning Phyiscs and Effects (Cambridge Univeristy Press, 2003).
7. D. J. Bocippio, S. J. Goodman and S. Heckman, *J. Appl. Meteor.* **39**, 2231 (1990).
8. M. Bernardi and D. Ferrari, *J. Electrostatics* **60**, 131 (2004); M. Bernardi and D. Ferrari, *Proc. 26th International Conference on Lightning Protection*, Krakow (Poland), 1, 100 (2002); M. Bernardi, C. Giorgi and V. Biscaglia, *Proc. 24th International Conference on Lightning Protection*, Birmingham-UK, 1, 187 (1998); R. Iorio and D. Ferrari, *Proc. 23rd International Conference on Lightning Protection*, Florence (Italy), 1, 191 (1996).
9. K. L. Cummins, M. J. Murphy, E. A. Bardo, W. L. Hiscox, R. B. Pyle and A. E. J. Pifer, *J. Geophys. Res.* **103**, 9035 (1998); E. P. Krider and R. C. Noggle, *J. Appl. Meteor.* **14**, 252 (1975).
10. R. E. Orville, R. A. Weisman, R. B. Pyle, R. W. Henderson and Jr. R. E. Orville, *J. Geophys. Res.* **92** (D5), 5640 (1987); U. Finke and T. Hauf, *Contributions to Atmos. Phys.* **69**, 361 (1996).
11. A. Papoulis, Probability, Random Variables, and Stochastic Processes (McGraw-Hill, New York, 1990).
12. S. Havlin, L.A. N. Amaral, Y. Ashkenazy, A. L. Goldberger, P. Ch. Ivanov, C.-K. Peng and H. E. Stanley, *Physica A* **274**, 99 (1999).
13. D. R. Cox and V. Isham, Point Processes (Chapman and Hall, London, 1980).
14. S. B. Lowen and M. C. Teich, *Fractals* **3**, 183 (1995).
15. R. G. Turcott, S. B. Lowen, E. Li, D. H. Johnson, C. Tsuchitani and M. C. Teich, *Biol. Cybern.* **70**, 209 (1994).
16. D. W. Allan, *Proc. IEEE* **54**, 221 (1966).
17. L. Telesca, V. Cuomo, V. Lapenna and M. Macchiato, *Geophys. Res. Lett.* **28**, 4323 (2001); L. Telesca, V. Cuomo, V. Lapenna and M. Macchiato, *Tectonophysics* **330**, 93 (2001); L. Telesca, V. Cuomo, V. Lapenna and M. Macchiato, *J. Seismology* **6**,

125 (2002); L. Telesca, V. Lapenna and M. Macchiato, *Phys. Earth Planet. Int.* **131**, 63 (2002); L. Telesca, V. Cuomo, V. Lapenna and M. Macchiato, *Fluctuation Noise Lett.* **2**, L357 (2002); L. Telesca, V. Cuomo, V. Lapenna and M. Macchiato, *Fractals* **10**, 67 (2002); L. Telesca, V. Lapenna and M. Macchiato, *Chaos Solit. & Fractals* **18**, 203 (2003); L. Telesca, V. Lapenna, M. Lovallo and M. Macchiato, *Chaos Solit. & Fractals* **21**, 387 (2004).

18. S. B. Lowen and M. C. Teich, *J. Acoust. Soc. Am.* **99**, 3585 (1996).
19. S. Thurner, S. B. Lowen, M. C. Feurstein, C. Heneghan, H. G. Feichtinger and M. C. Teich, *Fractals* **5**, 565 (1997).

A CORNUCOPIA OF CONNECTIONS: FINDING FOUR FAMILIAR FRACTALS IN THE TOWER OF HANOI

DANE R. CAMP

New Trier High School, 385 Winnetka Avenue
Winnetka, Illinois 60093 USA

This exposition draws a connection between the Tower of Hanoi puzzle and four well-known fractals. These connections are illustrated by examining a graph theoretical picture of the solution to the puzzle and identifying patterns within the structure of the graph. The numbering scheme introduced by Ian Stewart, in an article he wrote for Scientific American, connects a graphical representation of the solution to the Tower of Hanoi to the structure of Sierpinski's Gasket. Further investigation of the solution by the author shows how connections can also be discovered between the Tower of Hanoi and three other fractals: the Dragon Curve, Cantor Dust, and the Von Koch Curve.

1 The Towers of Hanoi

In a land far away in a long distant time
the priests in a temple played a game quite sublime.

With 100 disks that were set upon
3 stout upright pegs they played dusk until dawn.

For the priests of Hanoi knew when the game was done
the whole world would end and so would their fun.

The rules were quite simple for there were but two,
and if you listen closely, I'll share them with you.

The disks could be moved to pegs one at a time,
to put larger on smaller was considered a crime.

There were 100 priests, who all played this game
they each had one disk and took turns by name.

The 100th priest would utter a curse,
for he could not move, till the others moved first.

So the job was now left to priest 99,
who also cursed softly, and took place in line.

And so they all lined up cursing and recursing,
waiting their turn, and mentally rehearsing,

their one special move their own sacred place
in the cosmic scheme and the human race.

So let us attempt to act out this drama
for 4 little disks to lessen the trauma.

And when we are through with this smaller version,
you will then understand, what we call recursion.

The Tower of Hanoi is the stuff of legend. My students giggle when we start out our lesson on recursion with this poem and then act out the procedure using volunteers from the class. After the exercise, I always admit to them that the Tower of Hanoi puzzle is really not ancient; it was first marketed by French mathematician Edouard Lucas in 1883. However, I add that nesting puzzles of this type have a long and rich history, transcending cultural boundaries for they seem to have a deep broad based appeal to the human psyche. It is because of this almost mystical universality, triggered by my exposure to Ian Stewart's description of a graph theoretical approach to a solution to the Tower of Hanoi, and coupled with my intense passion for fractal geometry, that I decided to search for other familiar fractal forms within the puzzle's solution.

The puzzle typically consists of n disks that are stacked in descending order of size on a peg as shown below in Figure 1 for $n = 4$. The object is to move all of the disks from the present peg to one of the other empty pegs, using the additional peg as needed, while obeying the following rules:

1. Disks must only be moved one at a time from one peg to another.
2. A larger disk cannot be placed upon a smaller disk.
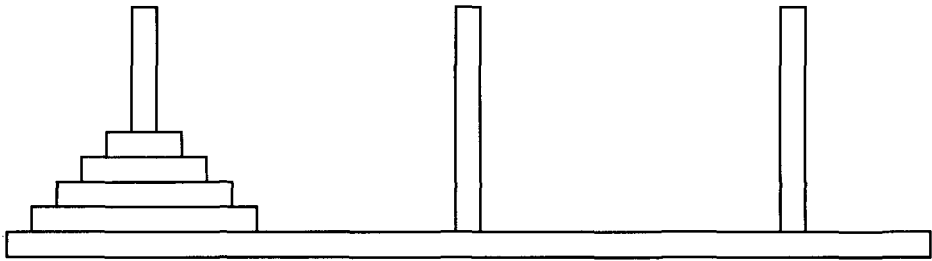
Figure 1. The Tower of Hanoi for four disks. (Comment: Experiment by simulating the puzzle with a quarter, nickel, penny and dime.)

2   The Tower of Hanoi and Sierpinski's Gasket

In an article about Sierpinski's Gasket, Ian Stewart describes a method for illustrating the solution to the Tower of Hanoi using graph theory. For one disk, assign the numbers 1, 2, and 3 to the states corresponding to which peg the disk is resting upon. Draw a graph representing the three states and connect those that are adjacent, meaning you can get from one to the other in one move, as in Figure 2.
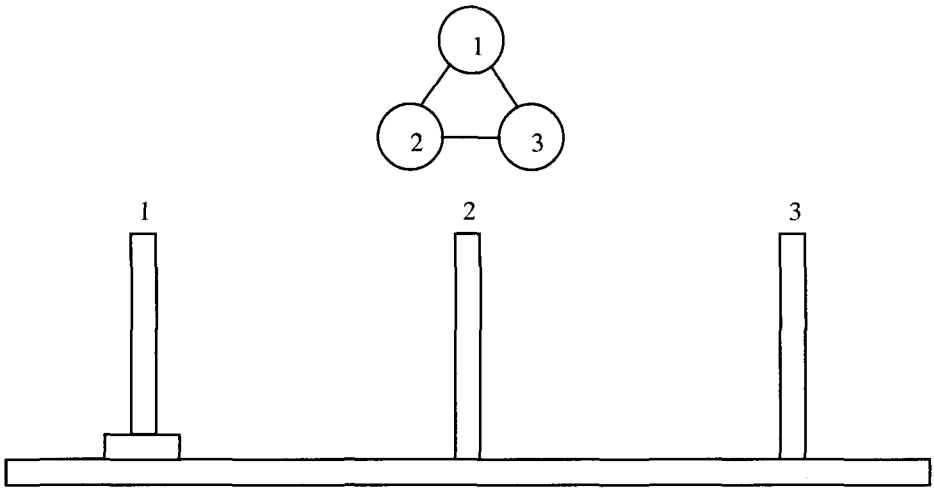
Figure 2. The Tower of Hanoi with one disk and its corresponding graph.

For two disks, do the same thing. In this case, the first [leftmost] digit represents the peg upon which the smallest disk rests. The second digit represents the location of the larger disk. Figure 3 shows that the graph contains all solutions to the puzzle for two disks, namely moving from one vertex to another along the side of the triangle. As a matter of fact, any path on the graph corresponds to a series of moves in the puzzle—even those that are not solutions.
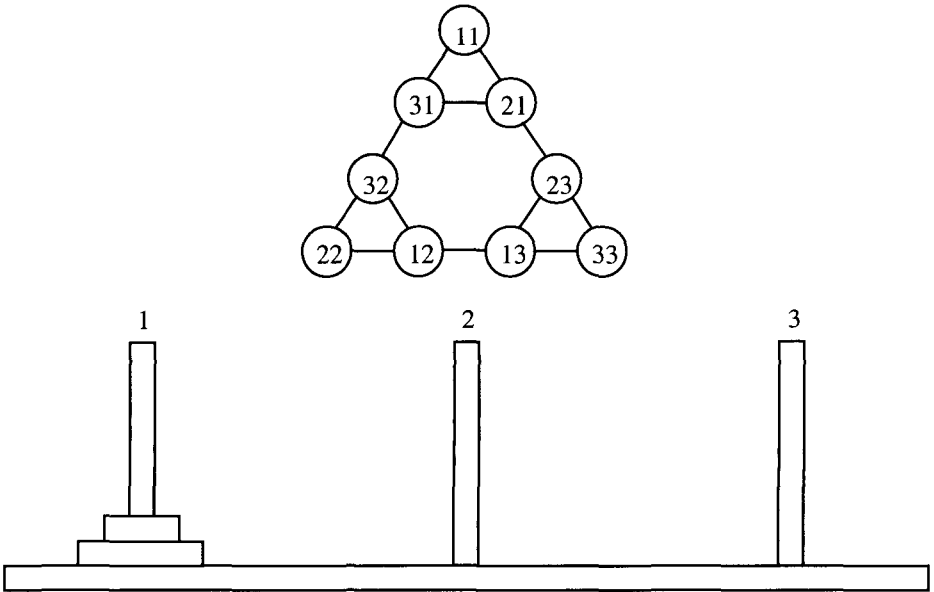
Figure 3. The Tower of Hanoi for two disks and its corresponding graph.

For three disks, the first digit represents the location of the smallest disk, the second corresponds to the medium sized disk and the last, to the largest disk. The solutions again are represented as paths from one vertex to another along the side of the triangle that is illustrated in Figure 4.
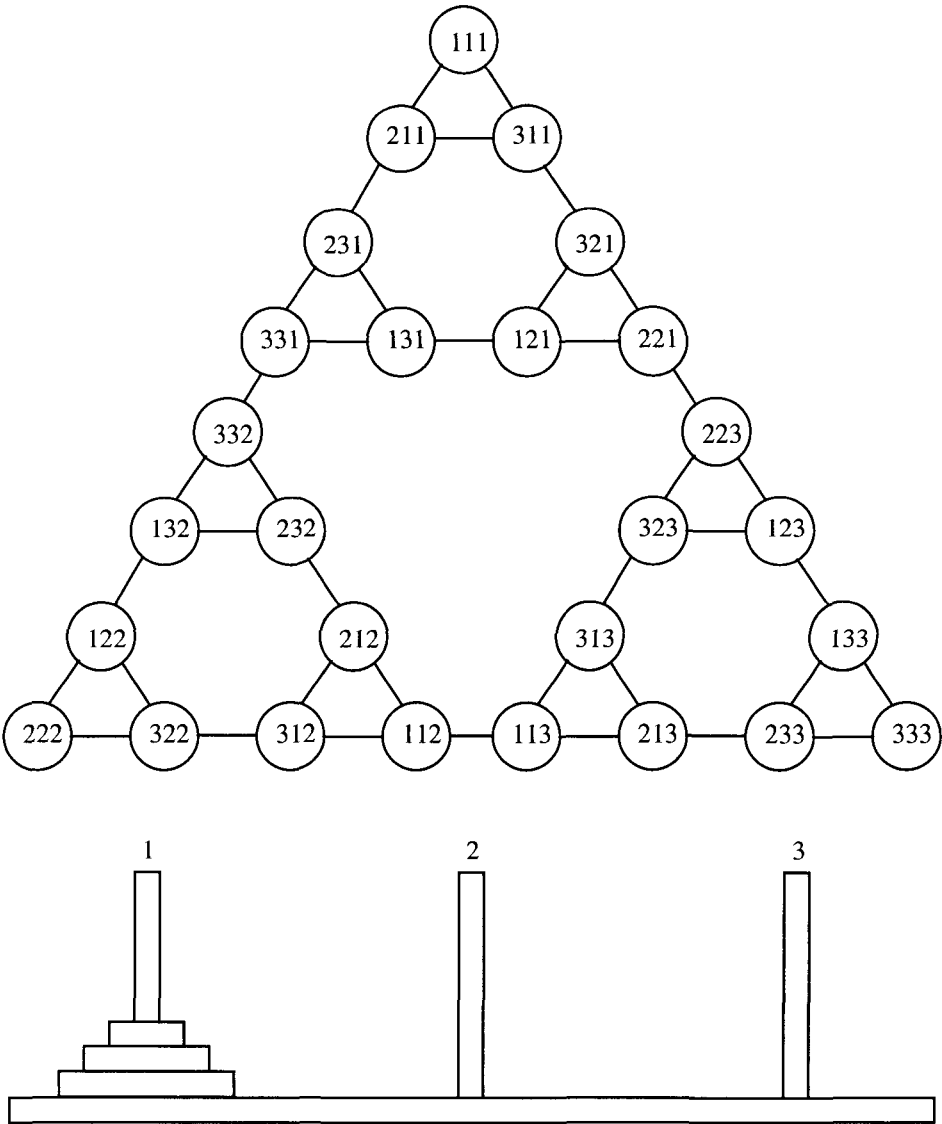
Figure 4. The Tower of Hanoi for three disks and its corresponding graph.

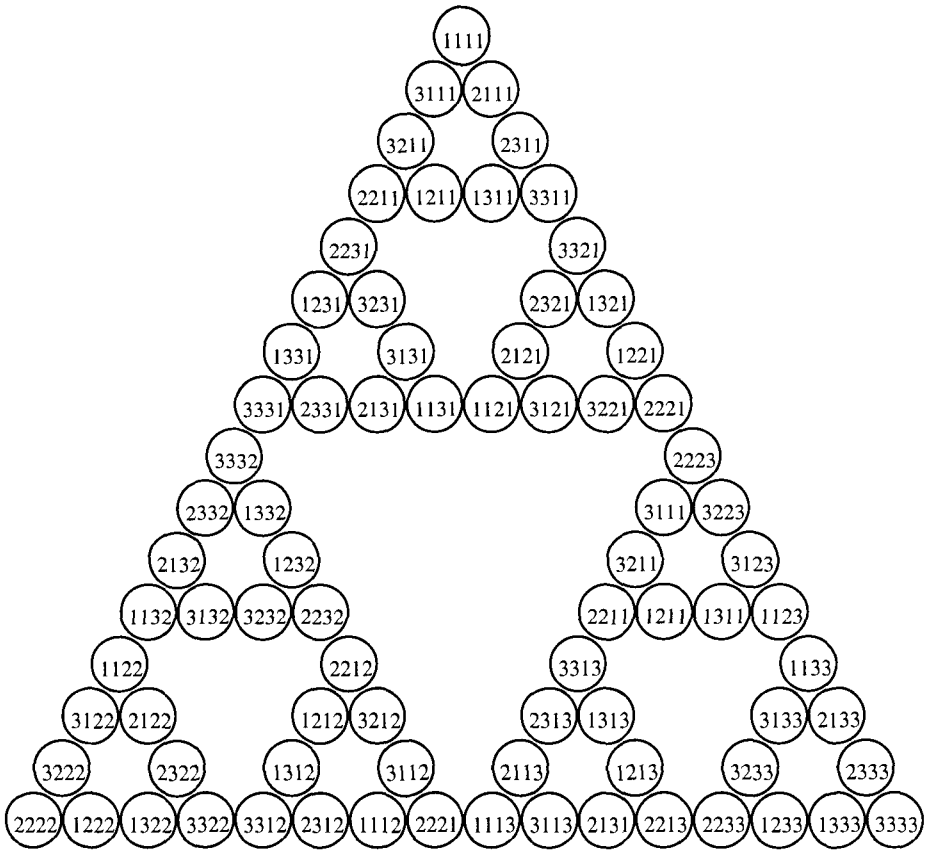Even before we look at the fourth iteration in figure 5, it is clear that the structure of the graph resembles that of the Sierpinski Gasket.

Figure 5. The graph corresponding to the solutions of the Tower of Hanoi for four disks. Again, the leftmost digit represents the peg upon which the smallest disk rests and the rightmost digit represents the peg upon which the largest disk rests. Notice all of the interesting numerical patterns!

3 The Dragon Curve

Not only do we have to wonder, in both senses of the word, about the appearance of Sierpinski's gasket in the solution to the Tower of Hanoi, but we should be curious as to whether other familiar fractals might also be connected to the Tower of Hanoi. Consider the stages of the Sierpinski Gasket solution to the Tower of Hanoi. Imagine that each disk in the original Tower has a head and a tail, like a coin. Each time you move a disk, you flip it over. Let L represent a fLip and R represent a flip back, Right side up, from tail to head. Now label the *edges* of the gasket in the following way:

1. Pick a starting vertex, for example the top.
2. Pick the target vertex, for example the bottom right.
3. As you progress from the initial vertex to the terminal vertex, label the edge with an L each time you would flip and an R each time you would put the disk right side up as illustrated along one side of the triangle in Figure 6.
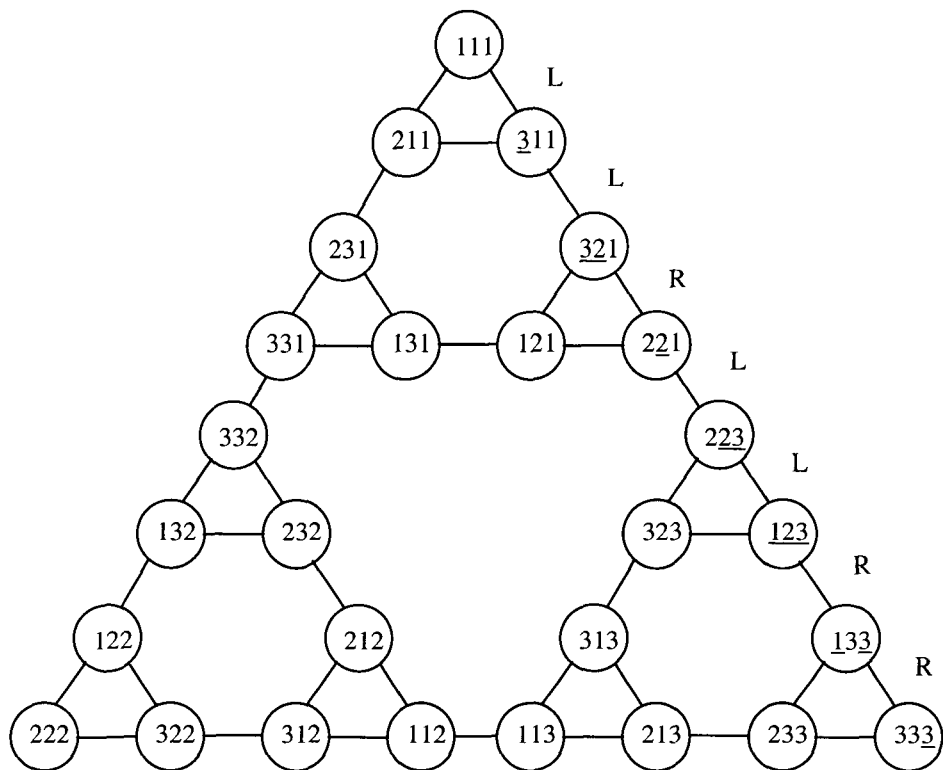
Figure 6. Coin flipping scheme for a solution to the Tower of Hanoi with three disks. Note that the coins, which are "tail up", are underlined.

Now, take the sequence of L's and R's that result from this procedure. Imagine that you are a turtle and the sequence represents directions for moving. Now draw the sequence by moving one unit then turning 90° Left or Right, then moving one unit turning again until the sequence is finished. You have just traced Figure 7, the third stage of a Dragon Curve.
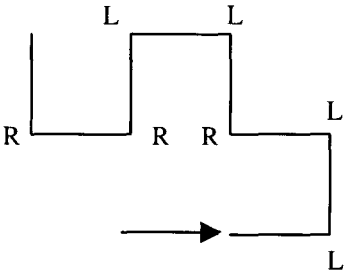

Figure 7. Converting the L's and R's to left and right turns yields the third iteration of the Dragon Curve.

4 Cantor Dust and the Dragon Curve

This time let's start with a fractal—Cantor Dust. The first stage of Cantor Dust is a line segment; let's label it L. Underneath the first stage, put the second stage. This is found by duplicating the stage above and removing the middle third. Call the left hand segment L and the right hand segment R. Continue constructing the Dust by duplicating the previous stage and removing the middle third of each segment labeling the segments with an L or an R depending upon whether they are to the left or right of the segment that you removed. Completing four stages will yield figure 8. Essentially, you now have a tree structure made up of L's and R's.
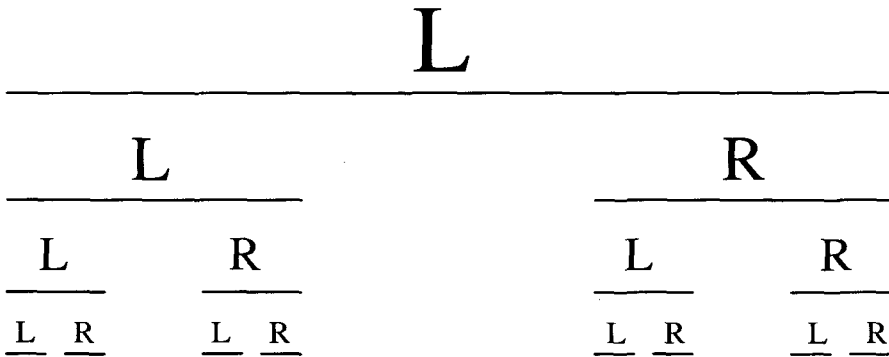

Figure 8. Stages of Cantor Dust organized in a tree fashion labeled as Left and Right branches.

Now perform an inorder traversal of the tree. Upon completing stage 4 you would have LLRLLRRLLLRRLRR, representing a 4th stage Dragon Curve corresponding to the edge of a 4th stage Sierpinski Gasket derived from the solution to the Tower of Hanoi with 4 disks! This is demonstrated in figure 9.
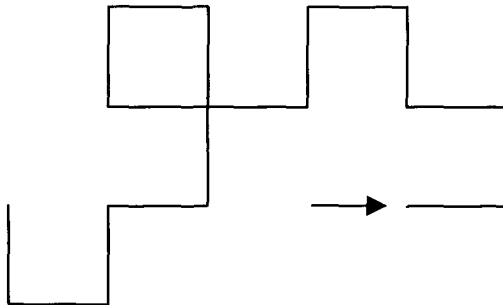

Figure 9. The fourth iteration of the Dragon Curve found by inorder traversal of the tree of L's and R's which was derived from generating Cantor Dust.

5 The Von Koch Curve

Finally, let us resort to an alternate coding scheme for the solution to the Tower of Hanoi. Let L represent a move to the left one unit and R represent a move to the right one unit, where moves "wrap around," (i.e. a move from 1 to 3 is L and a move from 3 to 1 is R) and the size of the letter corresponding to the size of the disk. Figure 10 reveals the letter sequence using this scheme for the Tower of Hanoi for 4 disks.

Figure 10. Right and left moves in a solution to the Tower of Hanoi with disk sizes highlighted for clarity.

So the solution for four disks could be represented as the sequence shown below in Figure 11:

$$\text{\tiny R}\text{\small L}\text{\tiny R}\text{\large R}\text{\tiny R}\text{\small L}\text{\tiny R}\text{\LARGE L}\text{\tiny R}\text{\small L}\text{\tiny R}\text{\large R}\text{\tiny R}\text{\small L}\text{\tiny R}$$

Figure 11. Right and left moves in a solution to the Tower of Hanoi for 4 disks.

As you can see, this is a different sequence than the one that yielded the Dragon Curve. However, a fractal still lurks inside if we are willing to be creative. Again, pretend that you are a turtle, but this time allow R to represent a 60 degree turn clockwise and L to represent a 120 degree turn counterclockwise. Does the resulting figure 12 look familiar? Yes, it resembles the Von Koch Curve. It should be noted, however, that this connection requires that there is an even number of disks in the original tower.
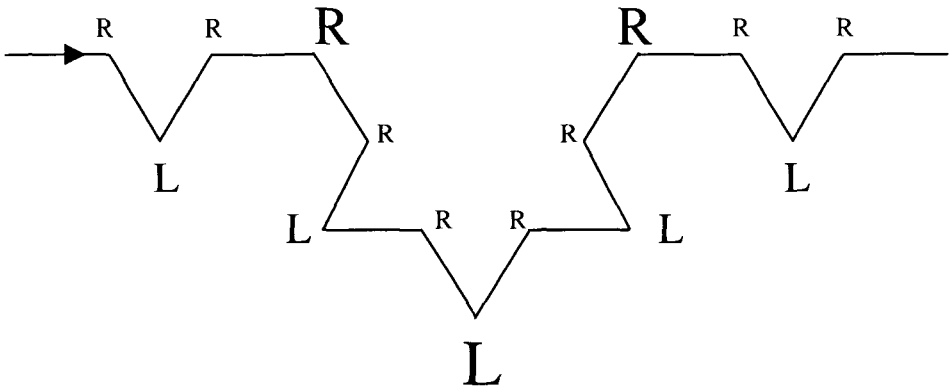
Figure 12. Translating the moves from Figure 10 into the second iteration of the Von Koch Curve.

6   Conclusion

As we have seen, the Tower of Hanoi is a rich source of patterns in general and fractals in particular. One can't help but wonder why such a wealth of connections naturally unfold from such a simple game. The occurrence of so many fractal structures within the solution of a puzzle that involves iteration should not be surprising, however, discovering them is still fascinating. These explorations clearly illustrate one of the apparently contradictory characteristics of fractal geometry that seems to be so pervasive—the feeling that once a fractal structure is identified within the context of a problem or a natural phenomenon, it is often both unexpected and obvious at the same time. As I searched for fractals within the solution to the Tower of Hanoi, I knew I would find them. I am also certain that there are even more.

Throughout this investigation I was reminded that mathematics is the language of patterns. It reinforced my conviction that it is important for me, as a high school teacher, to find ways to share the love of patterns with my students and also to instill in them a passion for identifying and discovering these patterns themselves. I also can't help but notice that whenever I am looking for patterns, I constantly find myself tripping over fractals. I hope that I never cease to experience genuine surprise and wonder when this happens.

332

References

1. Choate Jonathan, Robert Devaney, Alice Foster, <u>Iterations: A Tool Kit of Dynamics Activities</u>. Key Curriculum Press 1999, p. 101 and p. 173.
2. Dubrovsky, Vladimir, Nesting Puzzles Part II: Chinese rings produce a Chinese monster, *Quantum*, March/April 1996, pp. 61-65.
3. Maurer, Stephan B., and Anthony Ralston, *<u>Discrete Algorithmic Mathematics,</u>* A. K. Peters, Natick MA, 1998, pp. 102-108.
4. Stewart, Ian, Mathematical Recreations: Sierpinski's Ubiquitous Gasket, *Scientific American*, August 1999, pp. 90-91.
5. Stewart, Ian, A Fractal Guide to Tic-Tac-Toe, *Scientific American,* August 2000, pp. 86-88.
6. Vasilyev, Nikolay, and Victor Gutenmacher, Dragon Curves, Fear not! They're as tame as they are beautiful, *Quantum*, September/October 1995, pp.5-10.

MONITORING THE DEPTH OF ANAESTHESIA
USING FRACTAL COMPLEXITY METHOD *

W.KLONOWSKI, E.OLEJARCZYK, R.STEPIEN

Institute of Biocybernetics and Biomedical Engineering,
and GBAF, Medical Research Center,
Polish Academy of Sciences, Warsaw, wklon@ibib.waw.pl or gbaf@cmdik.pan.pl

P.JALOWIECKI , R.RUDNER

Chair and Department of Anaesthesiology, Intensive Therapy and Emergency Medicine,
Silesian University of Medicine, Sosnowiec, Poland

We propose a simple and effective method of characterizing complexity of EEG-signals for monitoring the depth of anaesthesia using Higuchi's fractal dimension method. We demonstrate that the proposed method may compete with the widely used BIS monitoring method.
Key-Words: - Complexity, Fractal Dimension, EEG, Anaesthesia.

1. Material and methods

We proposed a new method for monitoring the depth of anaesthesia by assessing fractal complexity of EEG-signal (cf. [1]). There exist several methods of measuring fractal dimension of EEG time series ([2],[3]) and there is an extensive literature on applying fractal measures from nonlinear dynamics to physiological signals ([4]-[6]). When using linear methods, like Fourier spectra or wavelet analysis, one forgets that these methods work properly only for stationary signals while biosignals are very non-stationary; similarly, when using time-delay method (cf. [7]) for embedding physiological time series in a multidimensional phase space and then calculating nonlinear characteristics such as e.g. correlation dimension, one forgets that 'embedology' works properly only for signals generated by point sources, while biosignals like EEG have complex spatio-temporal source structure [8]. To model EEG changes observed during general anesthesia that result from spatially-dispersed changes in electrocortical activity, one needs parallelized simulations of nonlinear partial differential equations on a grid representing an entire human cortex; it requires selection of plausible parameter sets from dozens of thousands possible sets and still there are dozens of such plausible sets [9]. The method proposed by Higuchi [10] yields the fractal dimension directly in time domain without necessity of embedding the data in a phase space; the algorithm (cf. Appendix 1) is very simple and so it may be applied in real time. It works even with relatively short epochs, also for nonstationary signals; it is also highly noise-resistant [11].

Following approval from the local Ethics Committee 58 patients during general anaesthesia were studied. Patients with known neurological disorders, as well as those who were receiving neuropsychiatric medication, were excluded from the study. Anaesthesia was induced with fentanyl, midazolam and propofol.

Brain electrical activity in all patients was measured continuously with an A-2000 BIS Monitor (software version: XP, Aspect Medical Systems, Newton, MA, USA) and bispectral index (BIS) was recorded every 10 seconds. The bispectral index is commonly accepted as a measure of hypnosis during anaesthesia [12], but the algorithms the BIS Monitor uses are not in public domain.

In order to calculate fractal dimension the unprocessed EEG signal was obtained using two different data acquisition systems.

In the first group (30 patients) - the raw, unprocessed EEG signal (±50 µV) was recorded from BIS Monitor onto PC with sampling rate 128 Hz using BSA program for BIS (version 3.22B2 for A-2000, programmed by S. Hagihira [13]).

In the second group (28 patients), EEG signal was registered by DigiTrack data acquisition system (ELMIKO, Warsaw, Poland), sampled with 250 Hz and filtered with a bandpass filter 3.5–50 Hz .

The recorded copy of the EEG was reviewed and proceed postoperatively to calculate Higuchi's fractal dimension. The results were averaged every 10 s for epochs 30 s long. Since $1 \leq D_f \leq 2$ the fractal dimension value has been presented as it's derivative: $(D_f - 1)*100$ to adjust the scale for better comparison with BIS.

In addition, depth of anaesthesia was continuously tested and classified by a specialist-anaesthesiologist to six OAA/S (Observer's Assessment of Alertness and Sedation) levels [14] (cf. Appendix 2); patients were judged to be conscious if the OAA/S score was between 3 - 5 and unconscious if the OAA/S score was less then 3. Pearson or Spearman correlation coefficient (r) were used to compare D_f with BIS Index.

2. Results

We analyzed EEG-signals post-operatively and demonstrated that the fractal dimension corresponds to the depth of anaesthesia. Below we demonstrate four interesting cases. In each case we show three variables - OAA/S, BIS, and fractal dimension $(D_f-1)*100$ during the whole procedure. In addition exact dosage and timing of drugs administration is illustrated.

Case nr 1
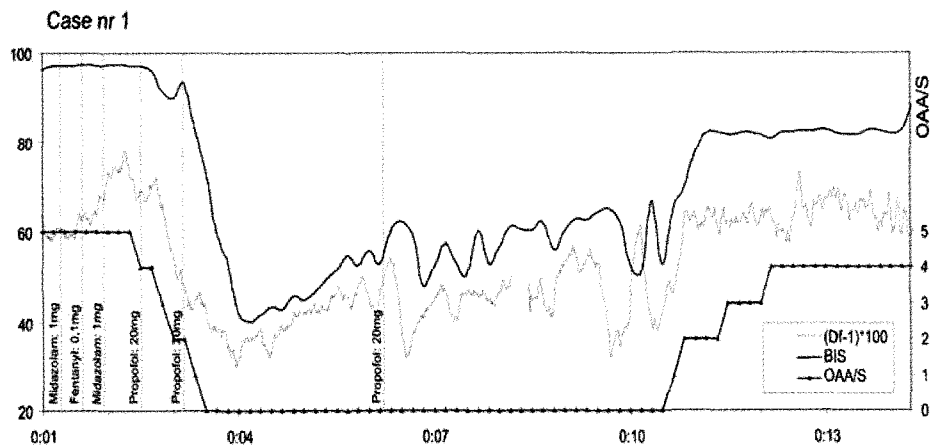
Case 1. A 40-year-old female, weighting 67 kg was anaesthetised for elective colonoscopy. Anaesthesia was induced with fentanyl (0,1 mg), midazolam (2 mg) and propofol (titrated to keep OAA/S bellow 2) up to a total dose of 60 mg in this case. The fractal dimension and the BIS present basically similar behaviour during induction, maintenance and termination of anaesthesia. But the $(D_f-1)*100$ started to decrease 25 seconds before the alteration of BIS value after the initial dose of propofol. Essential changes in the $(D_f-1)*100$ occurred 5 to 25 seconds earlier then BIS changes during the whole procedure.
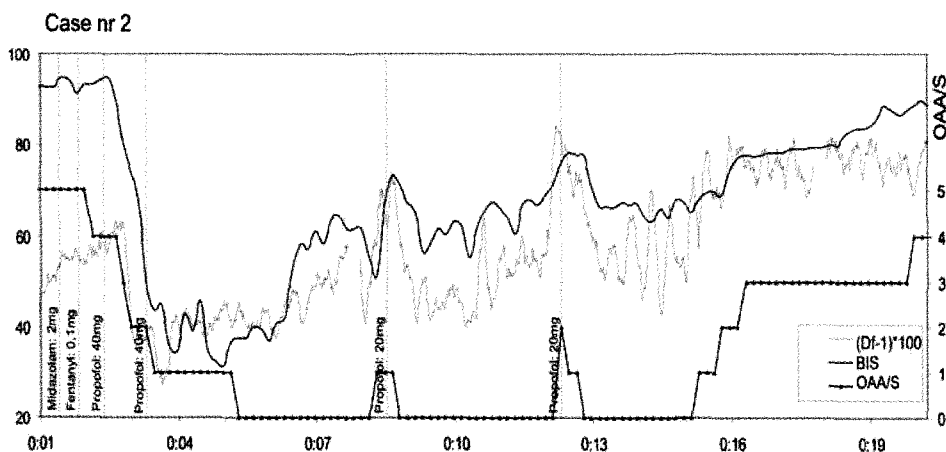
Case nr 2

Case 2. A 62-year-old female, weighting 68 kg, anaesthetized for elective colonoscopy in the similar way as in case nr 1. Anaesthesia was lightened twice during the procedure (OAA/S score greater than 0). At the moments of return of eye-lid reflex $(D_f-1)*100$ rose rapidly to rich higher level then before sedation. Those increases in $(D_f-1)*100$ were more significant and occurred earlier then rises of BIS. This finding could have predictable value to avoid unnecessary awakenings during anaesthesia.

336

Case nr 3
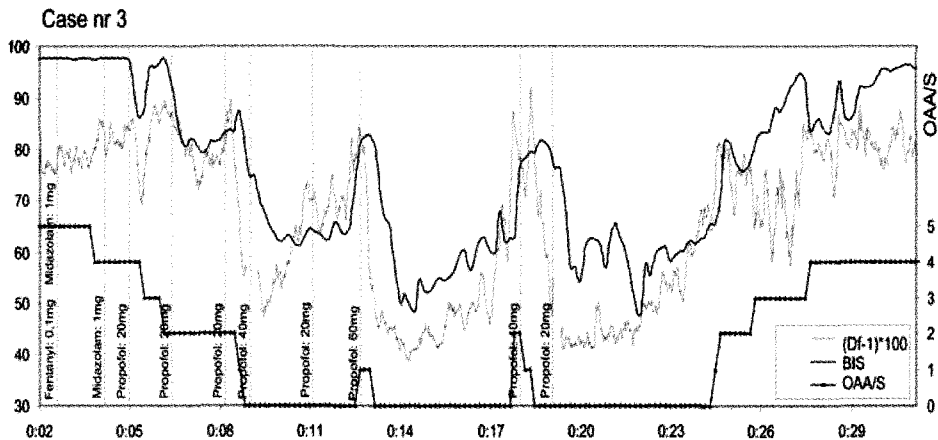
Case 3. A 31-year-old male, weighting 76 kg, was anaesthetized for colonoscopy as described above. In spite of administrating much higher doses of propofol, moments of lightening of anaesthesia were observed at 12th and 18th minutes of the study. Increasing OAA/S score evoked almost parallel changes in values of BIS and of $(D_f-1)*100$. Once again changes of $(D_f-1)*100$ occurred earlier and more rapidly than increasing of BIS value.

Case nr 4
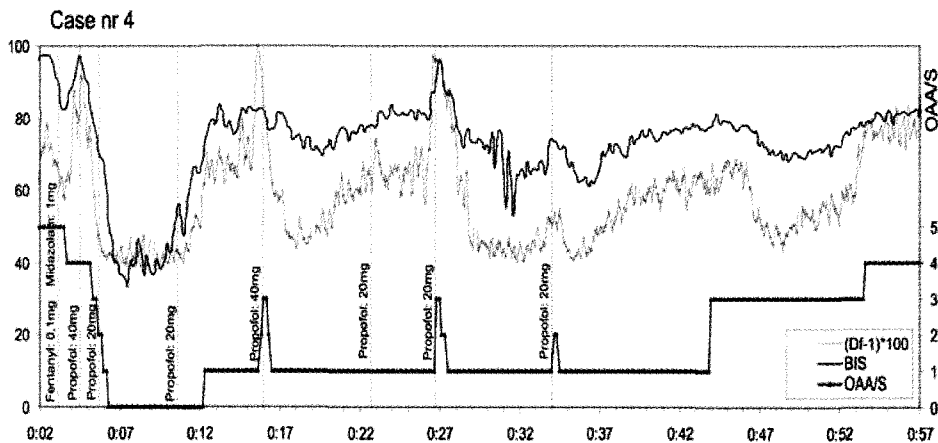
Case 4. In this case sedation (instead of anaesthesia) has been employed to secure adequate comfort for 65-year-old male patient during long, unpleasant procedure (colonoscopy with polypectomy). Sedation has been controlled according to the BIS (target BIS value between 60 and 80) with intermittent boluses of propofol. At 16th and 27th minutes of the study sedation was lightened to level 3 in OAA/S score. Record showed that at those moments $(D_f-1)*100$ rose rapidly to rich the highest value. BIS increased only once (at 27th minute) while at 16th minute remained unchanged. Awakening during and in the end of sedation can be predicted by rise of the $(D_f-1)*100$ towards its highest value.

Tables 1 and 2 show percentage of: almost perfect (1), very high (2), high (3), and average (4) positive correlation between D_f and BIS during anaesthesia (induction and awaking from anaesthesia).

On the results of the first group of 30 patients the Pearson parametric test of correlation between fractal dimension and BIS was performed. Fractal dimension values were averaged in 30-second window every 10 second to compare both measures correctly. Fractal dimension changes correlate well with BIS behaviour during anaesthetic induction (mean Pearson correlation coefficient r=0.81±0.2; p=0.02) and during awakening from anaesthesia (r = 0.76±0.2; p=0.016) (Table 1).

Parameters distributions in the second group of patients in most cases were not normal (Shapiro-Wilk test was performed to prove it). So, nonparametric Spearman test was performed in this case (Table 2).

Table 1. Percentage of positive correlation between D_f and BIS
during anaesthetic induction and during awakening from anaesthesia (1st group).

	r	induction – n (%)	awakening – n (%)
1	$0.9 \leq r < 1$	13 (43.3%)	9 (30%)
2	$0.7 \leq r < 0.9$	12 (40%)	12 (40%)
3	$0.5 \leq r < 0.7$	2 (6.7%)	5 (16.7%)
4	$0.3 \leq r < 0.5$	3 (10%)	4 (13.3%)

Table 2. Percentage of positive correlation between D_f and BIS
calculated for the whole procedure under anaesthesia (2nd group).

	r	n (%)
1	$0.9 \leq r < 1$	5 (17.86%)
2	$0.7 \leq r < 0.9$	18 (64.29%)
3	$0.5 \leq r < 0.7$	3 (10.71%)
4	$0.3 \leq r < 0.5$	2(7.14%)

Our result demonstrate that correlation between D_f and BIS in both groups of patients are very high (r > 0.7) in over 70% of cases.

338

3. Conclusions

There is a still growing need for relatively simple methods for computer-assisted non-invasive monitoring of patient in OR's and ICU's. Fractal methods of biosignal analysis can be very useful for these purposes. Application of methods borrowed from Nonlinear Dynamics and Symbolic Dynamics seems to be very promising – the methods are efficient and user-friendly.

Medicine may benefit from methods of time series analysis developed in Nonlinear Dynamics and Deterministic Chaos Theory, in particular from fractal and symbolic methods of analysis in time domain. As yet there is no consensus on the appropriate mathematical analysis of EEG-signals. Each method has its inherent advantages and disadvantages. Since a clinician must remain a necessary 'component' of diagnostic procedure he/she must understand the strength and the limitations of the computer-assisted assessment procedure. That is why, despite inherent nonlinearity and nonstationarity of EEG, applied methods must be sufficiently simple to be understandable for clinicians. It seems that our fractal method may fulfil these demands of relative simplicity.

The method we propose may find several applications. It may be used for analysis of other biosignals, like EOG, ECG, etc. Because it is simple and sufficiently quick to be applied in a real time it may be used with new nano-sensors. In general, application in biosignal analysis of the methods borrowed from Nonlinear Dynamics seems to be very promising – the methods are efficient for computer-assisted medical assessment. We have previously successfully applied Higuchi's method to different biosignals analysis, like automatic sleep staging based on analysis of sleep-EEG ([14], [15]). Here we have shown that Higuchi's fractal dimension method may be successfully applied to analysis of EEG-signal for monitoring of the depth of anaesthesia. The algorithm is easy and fast with respect to other methods and gives compression of information - on one-page it is possible to have information about EEG signal few hundred minutes long. The algorithm is also very quick and may be implemented in real time even on a PC. The patent for monitoring the depth of anaesthesia using fractal complexity method is pending [17].

Appendix 1

Higuchi's Fractal Dimension Method

Higuchi's algorithm measures *fractal dimension* of a time series directly in the time domain. It is based on a measure of length, $L(k)$, of the curve that represents the considered time series while using a segment of k samples as a unit. If $L(k)$ scales like

$$L\left(k\right) \sim k^{D_f}$$

the curve is said to show *fractal dimension* D_f. Because a simple curve has dimension equal 1 and a plane has dimension equal 2 value of D_f is always between 1 (for a simple curve) and 2 (for a curve which nearly fills out the whole plane). D_f measures *complexity* of the curve and so of the time series this curve represents on a graph.

From a given time series: $X(1)$, $X(2)$, ... , $X(N)$ the algorithm constructs k new time series, X^k_m :

$$X(m),\ X(m+k),\ X(m+2k),...,\ X(m\ +\ \mathrm{int}((N\text{-}m)/k) \times k)$$

for $m = 1, 2, ..., k$ where m – initial time, k – time interval, int(r) – integer part of a real number r.

For example, for $k=4$ and $N=1000$ the algorithm produces 4 time series:

$$\mathbf{X}^4_1 = X(1), X(5), X(9), ..., X(997)$$
$$\mathbf{X}^4_2 = X(2), X(6), X(10), ..., X(998)$$
$$\mathbf{X}^4_3 = X(3), X(7), X(11), ..., X(999)$$
$$\mathbf{X}^4_4 = X(4), X(8), X(12), ..., X(1000)$$

The 'length' $L_m(k)$ of each curve X^k_m is then calculated as:

$$L_m\left(k\right) = \frac{1}{k}\left[\left(\sum_{i=1}^{\mathrm{int}\left(\frac{N-m}{k}\right)} |X(m+i \times k) - X(m+(i-1)\times k)|\right) \frac{N-1}{\mathrm{int}\left(\frac{N-m}{k}\right) k}\right]$$

where N – the total number of samples.

$L_m(k)$ is not the 'length' in Euclidean sense, it represents the normalized sum of absolute values of differences in ordinates of pair of points distant k (with initial point m). The length of curve for the time interval k, $L(k)$, is calculated as the mean of the k values $L_m(k)$ for $m=1, 2,..., k$:

$$L(k) = \frac{\sum_{m=1}^{k} L_m(k)}{k} \tag{A.1}$$

The value of fractal dimension D_f is calculated by a least-squares linear best-fitting procedure as the angular coefficient of the linear regression of the log-log graph of (A.1):

$$y = ax + \bar{b} \tag{A.2}$$

namely

$$\ln(L(k)) = D_f \ln\left(\frac{1}{k}\right) + \left(\bar{b} + \ln(c)\right)$$

according to the following formulae:

$$D_f = \frac{n\sum(x_k \, y_k) - \sum x_k \sum y_k}{n\sum(x_k^2) - \left(\sum x_k\right)^2}$$

where: $y_k = \ln(k)$, $x_k = \ln\left(\frac{1}{k}\right)$, $k = k_1, \dots, k_{max}$ and n denotes the number of k values for which the linear regression is calculated $(2 \le n \le k_{max})$.; c is a scaling factor - Higuchi's fractal dimension shows a scaling feature - multiplication of all amplitudes X^k_m by a constant factor, c, does not change D_f. The standard deviation of D_f is calculated as:

$$S_{D_f} = \sqrt{\frac{n\left[\sum y_k^2 - D_f \sum x_k y_k - \bar{b} \sum y_k\right]}{(n-2)\left[n\sum x_k^2 - \left(\sum x_k\right)^2\right]}}$$

where (cf. eq. (A.2)):

$$\bar{b} = \frac{1}{n}\left(\sum y_k - D_f \sum x_k\right)$$

with standard deviation

$$S_{\bar{b}} = \sqrt{\frac{1}{n} \times S_{D_f}^2 \times \sum x_k^2}$$

Appendix 2

Observer's Assessment of Alertness and Sedation Scale (OAA/S)

Responsiveness	Speech	Facial Expression	Eyes	Composite Score
Responds readily to name spoken in normal tone	Normal	Normal	Clear, no ptosis	5 (alert)
Lethargic response to name spoken in normal tone	Mild slowing or thickening	Mild relaxation	Glazed or mild ptosis (less than half the eye)	4
Responds only after name is called loudly or repeatedly	Slurring or prominent slowing	Marked relaxation (slack jaw)	Glazed and marked ptosis (half the eye or more)	3
Responds only after mild prodding or shaking	Few recognizable words			2
Does not respond to mild prodding or shaking				1
Does not respond to noxious stimuli. No eyelid reflex				0

References

1. R. Rudner, P. Jalowiecki, M. Willand, W. Klonowski, E. Olejarczyk, R. Stepien, S. Hagihira: Fractal dimension - a new EEG-based method of assessing depth of anaesthesia in comparison with BIS during induction and recovery from anaesthesia. *Euroanaesthesia 2005,* Vienna, Austria, May 28-31, 2005, *European Journal of Anaesthesiology* **22**, Suppl. 34, 2005, A-118,. pp.32-33.

2.   D. A. Chernik, D. Gillings, and H. Laine: Validity and reliability of the observer's assessment of alertness/sedation scale: Study with intravenous midazolam. *J. Clin. Psychopharm.* **10**, 244-245 (1990).

3.   J.B. Bassingthwaighte, L. Liebovitch, B.J. West: *Fractal Physiology.* Oxford University Press, 1994.

4.   B.J. West, M.N. Novaes, V. Kavcic: Fractal Probability Density and EEG/ERP Time Series. In: *Fractal Geometry in Biological Systems: An Analytical Approach* P.M. Iannaccone and M. Khokha (Eds.), pp. 267-316, CRC-Press, Boca Raton, 1996.

5.   I. Dvorak and J. Siska: On some problems encountered in the estimation of the correlation dimension of the EEG. *Physics Letters A*, **118** (2), p. 63-66 (1986).

6.   M. Palus, I. Dvorak, I. David: Spatio-temporal dynamics of human EEG. *Physica A* **185** (1-4), pp. 433-438 (1992).

7.   A.L. Goldberger, L.A.N. Amaral, J.M.Hausdorff, P.Ch. Ivanov, C.-K. Peng and H.E. Stanley: Fractal dynamics in physiology: Alterations with disease and aging. *PNAS* **99,** Suppl.1, pp. 2466-2472 (2002).

8.   Ying-Cheng Lay and N. Ye: Recent development in chaotic time series analysis. *Int. J. Bifurcation and Chaos* 13(6), pp. 1383-1422 (2003).

9.   N.G. Makarenko: Time series from geometry and topology of spatio-temporal chaos. *Izvestiya VUZ Applied Nonlinear Dynamics,* **12** (6), pp. 3-16 (Saratov , 2004, in Russian).

10.   I. Bojak and   D.T. Liley: Modeling the effects of anesthesia on the electroencephalogram. Physical Review E 71 (4), (2005).

11.   J. I. Rampil: A Primer for EEG signal processing in anaesthesia. *Anaesthesiology* **89**, 980-1002 (1998).

12.   T. Higuchi: Approach to an irregular time series on the basis of the fractal theory, *Physica D* **31**, 277-283 (1988).

13.   S. Hagihira, M. Takasina, T. Mori, T. Mashimo, I. Yashiya: Practical issues in bispectral analysis of electroencephalographic signals. *Anesthesia and Analgesia*; 93, 966-70, (2001)

14.   E. Olejarczyk: *Analysis of EEG signals using fractal dimension.* PhD Thesis, Supervisor Prof. W.Klonowski, Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, 2003.

15.   W. Klonowski, E .Olejarczyk, and R. Stepien: New Methods of Nonlinear and Symbolic Dynamics in Sleep EEG-Signal Analysis. In: *Modelling and Control in Biomedical Systems* (D. D. Feng and E. R. Carson, Eds.), pp. 241-244, IFAC Publications, Elsevier, Oxford, 2003.

16.   W. Klonowski, E. Olejarczyk, and R. Stepien: Complexity of Polysomnogram Signals. *WSEAS Transactions on Computers* 3(5), 1290-1294 (2004).

17.   W. Klonowski, E. Olejarczyk, R. Stepien, P. Jałowiecki, R. Rudner: Methods of Monitoring of the Depth of Anaesthesia, P-372355, Patent Office, Warsaw 2005.

THE COMPLEX COUPLINGS AND GOMPERTZIAN DYNAMICS

PRZEMYSLAW WALISZEWSKI

Department of Theoretical Chemistry, University of Poznan, Grunwaldzka 6
Poznan, PL 60780, Poland

JERZY KONARSKI

Department of Theoretical Chemistry, University of Poznan, Grunwaldzka 6
Poznan, PL 60780, Poland

The universal dynamics of growth and self-organization of cellular dynamic systems is described by the Gompertz function,[1] Eq. (1):

$$f(t) = e^{a(1-e^{-bt})} \tag{1}$$

in which $f(t)$ stands for example for a number of cells, volume, or their weight after time t, a and b are experimental coefficients determining slope of the curve. Since Gompertzian dynamics is so universal in nature, a question arises how this dynamics emerges?

For the purpose of this study, space is defined by a system of the geometrical co-ordinates. Those co-ordinates build up a volume, in which the nonlinear process with Gompertzian dynamics occurs. Time is a scalar with the same features as in any other dynamic equation. Owing to the relationship with the system of co-ordinates, time is a parameter and takes the sense of the evolutional co-ordinate.

First, we demonstrate that time and space are coupled to each other through nonlinear Gompertzian dynamics. This coupling determines the existence of fractal time-space in which growth occurs. The appropriate equation possesses the following algebraic form, Eq. (2):

$$\ln x = \frac{1}{\sigma - \sigma_0} \ln \frac{A_0 a}{A} + \frac{\beta}{\sigma - \sigma_0} \ln t \tag{2}$$

in which x is a scalar, geometrical variable, which locates an effect in space, t stands for scalar time, V_0 and V denote the volumes of the growing cellular system after time t_0 and t, respectively, A, A_0 stand for the scaling coefficients, a and b are experimental coefficients determining slope of the Gompertz curve, β is the temporal fractal dimension, σ is the spatial fractal dimension[1].

Second, we find that Gompertzian dynamics emerges at the macroscopic level as a result of coupling of stochastic phenomena at the microscopic level. Indeed, analysis of the linear differential equation of the first order, which generates the Gompertz function

as a solution, reveals the existence of a complex coupling of probabilities. If the normalized Gompertz function $f_N(t)$ is given by Eq. (3):

$$f_N(t) = e^{-e^{-bt}} \qquad (3)$$

and $b = r$, $f_N(t) = p(t)$, then $0 \le p(t) \le 1$ for any t, and $e^{-bt} = -\ln p(t)$, the normalized derivative gets a form given by Eq. (4):

$$p'(t) = rp(t)(-\ln p(t)) \qquad (4)$$

The Eq. (4) can be transformed to the iterated form given by Eq. (5):

$$p_{n+1} = rp_n(-\ln p_n) + p_n \qquad (5)$$

The iterated equation with the above-defined coupling of probabilities generates Feigenbaum-like diagram. That latter diagram indicates that the Gompertz function describes an equilibrium between regular processes, that is, processes with dynamics that are predictable for any time-point, (e.g., sinusoidal glycolytic oscillations) and chaotic processes, that is, processes with dynamics that are unpredictable in time, but are characterized by certain regularities, (e.g., the existence of strange attractor for any biochemical reaction)[1].

It is worth to notice that Gompertzian dynamics also defines the function of the Shannon entropy S:

$$f_N(t) = p(t) = \int_0^\infty p'(t)dt = -r\int_0^\infty p(t)\ln p(t)dt = rS \qquad (6)$$

This suggests the existence of a relationship between both fractal dimensions, Gompertzian dynamics and Kolmogorov-Sinai entropy (see Eq. 2).

Finally, we find that Gompertzian dynamics appears in the models with the simpliest couplings, such as the Langevin equation with the stationary differential Markov chain. Indeed, the conditional probability density $P(v, t)$ can be expressed in the form of the following equation with the Gompertz function $f(t)$ in its structure, Eq. (7):

$$P(v,t) = \int_{-\infty}^\infty \frac{dk}{2\pi}(e^{(-ik\int_0^{M(t)} e^{-\lambda(t-\tau)}dM(\tau))})f(t)|k|^\alpha \qquad (7)$$

Fractal Gompertzian dynamics is generated owing to both the complex coupling of probabilities of stochastic processes at the microscopic scale and the existence of fractal structure of time and space; the essence of life. The existence of fractal structure of time-space for non-linear phenomena with Gompertzian dynamics implies application of fractional differential equations in the mathematical models of growth.

Reference

1.   P. Waliszewski and J. Konarski, *Biosystems.* **80**, 91 (2005).

Index

Complexus Mundi
Emergent Patterns in Nature

The dynamics of complex systems can clarify the creation of structures in Nature. This creation is driven by the collective interaction of constitutive elements of the system. Such interactions are frequently nonlinear and are directly responsible for the lack of prediction in the evolution process. The self-organization accompanying these processes occurs all around us and is constantly being rediscovered, under the guise of a new jargon, in apparently unrelated disciplines.

This volume offers unique perspectives on aspects of fractals and complexity and, through the examination of complementary techniques, provides a unifying thread in this multidisciplinary endeavour. Do nonlinear interactions play a role in the complexity management of socio-econo-political systems? Is it possible to extract the global properties of genetic regulatory networks without knowing the details of individual genes? What can one learn by transplanting the self-organization effects known in laser processes to the study of emotions? What can the change in the level of complexity tell us about the physiological state of the organism? The reader will enjoy finding the answers to these questions and many more in this book.

Cover graphics courtesy of Nicoletta Sala, University of Mendrisio, Switzerland